THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Intelligence Artificielle

Présentée et soutenue par :

M. RICHARD ALLIGIER

le jeudi 13 novembre 2014

Titre :

APPRENTISSAGE ARTIFICIEL APPLIQUE A LA PREVISION DE TRAJECTOIRE D'AVION

Ecole doctorale :

Mathématiques, Informatique, Télécommunications de Toulouse (MITT)

Unité de recherche :

Institut de Recherche en Informatique de Toulouse (I.R.I.T.)

Directeur(s) de Thèse :

M. NICOLAS DURAND

M. DAVID GIANAZZA

Rapporteurs :

M. ERIC FERON, GEORGIA INSTITUTE OF TECHNOLOGY

M. MARC SCHOENAUER, UNIVERSITE PARIS 11

Membre(s) du jury :

M. MARC SCHOENAUER, UNIVERSITE PARIS 11, Président

M. DAVID GIANAZZA, DIR DES SERVICES DE NAVIGATION AERIENNE, Membre

M. HARTMUT FRICKE, TECHNISCHE UNIVERSITAT DRESDEN, Membre

M. MATHIEU SERRURIER, UNIVERSITE TOULOUSE 3, Membre

M. NICOLAS DURAND, ECOLE NATIONALE DE L'AVIATION CIVILE, Membre

Résumé

L'organisme Eurocontrol prévoit une forte hausse du trafic aérien européen d'ici l'année 2035. Cette hausse de trafic justifie le développement de nouveaux concepts et outils pour pouvoir assurer les services dus aux usagers de l'espace aérien. La prévision de trajectoires d'avion est au cœur de ces évolutions. Parmi ces outils, les outils de détection et résolution de conflits utilisent les trajectoires prédites pour anticiper les pertes de séparation entre avions et proposer des solutions aux contrôleurs aériens. L'horizon de prédiction utilisé pour cette application est de l'ordre de dix à vingt minutes.

Parmi les algorithmes réalisant une détection et résolution de conflits, certains sont mis en œuvre au sol, obligeant ainsi les prédictions à être calculées en n'utilisant que les informations disponibles dans les systèmes sols. Dans ces systèmes, la masse des avions ainsi que les profils de vitesse ou de poussée des moteurs ne sont pas connus. Ainsi, le calcul d'une trajectoire prédite avec un modèle physique se fait en utilisant des valeurs de référence pour les paramètres inconnus.

Dans ce cadre, nous nous intéressons à la phase de montée pour laquelle ces paramètres influent grandement sur la trajectoire de l'avion. Ce travail s'appuie sur le modèle physique Base of Aircraft DAta (BADA) développé et maintenu par Eurocontrol. Ce modèle physique modélise, entre autres, les performances des avions. Il fournit également des valeurs de référence pour les paramètres inconnus comme la masse de l'avion, son profil de vitesse en montée, ou la commande de poussée des moteurs. Ce modèle, largement utilisé dans le monde entier, est particulièrement imprécis pour la phase de montée, car les valeurs réelles de ces paramètres sont parfois très éloignées des valeurs de référence.

Dans cette thèse, nous proposons soit d'estimer directement certains paramètres, comme la masse, à partir des points passés de la trajectoire, soit d'utiliser des méthodes d'apprentissage supervisé afin d'apprendre, à partir d'exemples, des modèles prédisant les valeurs des paramètres manquants (masse, loi de poussée, vitesses cibles).

Ces différentes méthodes sont testées sur des données radar Mode-C et Mode-S sur plusieurs types d'avions. Les prédictions obtenues avec ces méthodes sont comparées à celles obtenues avec les paramètres de référence. Elles sont également comparées avec les prédictions obtenues par des méthodes de régression prédisant directement l'altitude de l'avion plutôt que les paramètres du modèle physique. Nos méthodes permettent de réduire, suivant le type de l'avion, de 50 % à 85 % par rapport à la méthode BADA de référence, la racine de l'erreur quadratique moyenne sur l'altitude prédite à un horizon de dix minutes

Abstract

The Eurocontrol organization forecasts a strong increase of the European air traffic till the year 2035. This growth justifies the development of new concepts and tools in order to ensure services to airspace users. Trajectory prediction is at the core of these developments. Among these tools, conflict detection and resolution tools use trajectory predictions to anticipate losses of separation between aircraft and propose solutions to air traffic controllers. For such applications, the time horizon of the prediction is about ten to twenty minutes.

Among conflict detection and resolution algorithms, some are operated in ground-based systems. The trajectory predictions must then be computed using only the information that is available to ground systems. In these systems, the mass, the speed profile and the thrust setting are unknown. Thus, using a physical model, the trajectory predictions are computed using reference values for unknown parameters.

In this context, we are focusing on the climb phase. In this phase these unknown parameters have a great influence on the aircraft trajectory. This work relies on a physical model of the aircraft performances : BADA, developed and maintained by Eurocontrol. It also provides reference values for unknown parameters such as the mass, the speed profile and the thrust setting. This widely used model is particularly inaccurate for the climb phase as the actual values for the unknown parameters might be very different from the reference values.

In this thesis, we propose to estimate directly the mass, an unknown parameter, using a physical model and past points of the trajectory. We also use supervised learning methods in order to learn, from examples, some models predicting the unknown parameters (mass, speed profile and thrust setting).

These different approaches are tested using Mode-C Radar data and Mode-S Radar data with different aircraft types. The obtained predictions are compared with the ones obtained with the BADA reference values. These predictions are also compared with predictions obtained by directly predicting the future altitude instead of the unknown parameters of the physical model. These methods, depending on the aircraft type, reduces the root mean square error on the predicted altitude at a 10 min horizon by 50 % to 85 % when compared to the root mean square error obtained using BADA with the reference values.

Remerciements

Un grand merci à David Gianazza et Nicolas Durand qui m'ont supporté lors de mes nombreux doutes divers et variés. La patience dont ils ont fait preuve a rendu moins pénible l'interminable rédaction du présent pavé. Je tiens à remercier David Gianazza, encore, pour sa bienveillance et pour m'avoir laissé la liberté d'explorer une approche plus physique du problème considéré dans cette thèse. Son aide m'a également été précieuse lors de la rédaction d'articles en langue étrangère.

Je souhaite remercier Marc Schoenauer et Éric Féron qui ont accepté d'être rapporteurs ainsi que les autres membres du jury, Hartmut Fricke et Mathieu Serrurier.

Je remercie Mohammad Ghasemi-Hamed pour sa simplicité et ses discussions parfois étonnantes et dans tout les cas toujours amusantes. Grâce à lui, je suis capable de tenir une conversation téléphonique en arabe. Je remercie également Charlie Vanaret pour ses jeux de mots dignes d'un vendredi, même un lundi. Il m'a souvent fait briller les yeux avec son jouet [1]. Je souhaite également remercier Cyril Allignol pour sa gentillesse, son écoute et ses jeux de mots qui font concurrence à ceux de Charlie ; Nicolas Barnier pour son fair-play à la pétanque et l'étendue de ses connaissances qu'il sait si bien partager ; Alexandre Gondran pour ses discussions politico-économiques toujours intéressantes et Jean-Baptiste Gotteland pour sa gentillesse, sa bonne humeur et la redécouverte de l'ADA chaque année en automne.

Merci à tous ceux qui m'ont aidé à exploiter ces ingrates données : François Huchet, Serge Roux, Jean-Paul Imbert, Bernard Brémond et Alain Hérout.

J'exprime toute ma reconnaissance aux personnels administratifs et techniques dont la disponibilité et la gentillesse ont rendu facile l'exercice au quotidien : Catherine Migot, Colette Roy, Sabine Cantayre, Serge Roux et Jean-Paul Imbert.

Enfin, je remercie ma famille : mes parents et mon frère qui m'ont toujours soutenu aussi loin que je me souvienne et même avant.

1. Son sujet de thèse et ce qu'il en a fait est très intéressant.

Table des matières

1 Contexte **7**
- 1.1 Gestion du trafic aérien . . . 7
  - 1.1.1 Différents types de vols . . . 8
  - 1.1.2 Différentes classes d'espace aérien . . . 8
- 1.2 Contrôle du trafic aérien . . . 10
  - 1.2.1 Vocabulaire et unités de mesure . . . 10
  - 1.2.2 Méthodes du contrôle . . . 11
- 1.3 Évolutions dans la gestion du trafic aérien . . . 12
  - 1.3.1 Évolution du trafic aérien en Europe . . . 12
  - 1.3.2 Évolution des concepts de gestion du trafic aérien . . . 12
- 1.4 Prévision de trajectoires et enjeux associés pour l'ATM . . . 13
  - 1.4.1 Méthodes existantes . . . 14
  - 1.4.2 Détection et résolution de conflits . . . 16
- 1.5 Cadre de travail . . . 17

2 Modélisation physique **19**
- 2.1 Plusieurs définitions d'altitudes . . . 20
- 2.2 Modélisation de l'atmosphère . . . 21
- 2.3 Application de la seconde loi de Newton . . . 22
- 2.4 Modèle avion BADA . . . 24
  - 2.4.1 Modélisation de la poussée . . . 24
  - 2.4.2 Modélisation de la consommation . . . 25
  - 2.4.3 Modélisation des forces aérodynamiques . . . 25
  - 2.4.4 *Energy Share Factor* . . . 26
  - 2.4.5 Ajustement des coefficients . . . 26
  - 2.4.6 Système d'équations du modèle avion BADA, avec prise en compte du vent . . . 27
- 2.5 Modélisation BADA des procédures compagnies . . . 28
  - 2.5.1 Profil de poussée pour la montée . . . 28
  - 2.5.2 Profil de vitesse air pour la montée . . . 29
- 2.6 Calcul de la trajectoire prédite . . . 29
  - 2.6.1 Calcul de la variation d'énergie spécifique . . . 30
  - 2.6.2 Calcul de l'excès de puissance spécifique . . . 31

2.6.3 Calcul du taux de montée et de la vitesse sol . . . 32
2.6.4 Résolution numérique du système d'équations différentielles . . . 33
2.6.5 Importance de l'excès de puissance spécifique . . . 34

3 Apprentissage supervisé et méthodes de régression 35
3.1 Apprentissage supervisé . . . 36
3.1.1 Notions générales sur l'apprentissage supervisé . . . 36
3.1.2 Évaluation des performances . . . 40
3.1.3 Choix des hyper-paramètres et ensemble de test . . . 41
3.1.4 Réduction du nombre de variables explicatives . . . 42
3.2 Méthodes de régression utilisées . . . 46
3.2.1 Régression linéaire . . . 46
3.2.2 Réseaux de neurones artificiels . . . 52
3.2.3 Gradient Boosting Machine . . . 54

4 Données de trajectoires 59
4.1 Obtention des exemples . . . 59
4.1.1 Types de données . . . 59
4.1.2 Deux jeux de données . . . 61
4.1.3 Lissage des données . . . 62
4.1.4 Filtrage des trajectoires . . . 72
4.1.5 Construction des exemples . . . 72
4.2 Analyse des exemples . . . 72
4.2.1 Profils de température observés . . . 74
4.2.2 Profils de vitesse observés . . . 74
4.2.3 Étude de la variation d'énergie spécifique . . . 77
4.2.4 Profils d'Energy Share Factor . . . 84
4.3 Conclusion . . . 87

5 Estimation de la masse à partir des points passés 91
5.1 Méthode naïve . . . 92
5.2 Méthode adaptative . . . 95
5.2.1 Ajustement de la masse . . . 95
5.2.2 Adaptation du paramètre de sensibilité . . . 96
5.3 Méthode des moindres carrés . . . 97
5.3.1 Prise en compte de la consommation : relation entre les m_i . . . 98
5.3.2 Approximation de $\mathcal{E}$ . . . 99
5.3.3 Minimisation de $\mathcal{E}_{approx}$ . . . 100
5.4 Méthode robuste . . . 100
5.4.1 Minimisation de $\mathcal{E}_{robuste}$ . . . 101
5.4.2 Newton par intervalles . . . 101
5.5 Comparaison sur des données simulées . . . 101
5.5.1 Ajout du bruit sur les grandeurs observées . . . 103

5.5.2 Résultats . . . 103
5.5.3 Influence de ΔT sur l'erreur . . . 107
5.5.4 Discussion sur les méthodes . . . 109
5.5.5 Limites et conclusion de cette étude . . . 109
5.6 Comparaison sur des données réelles . . . 110
5.6.1 Comparaison de la prédiction . . . 111
5.6.2 Distribution de la masse estimée sur les points passés . . . 112
5.6.3 Discussion sur la masse estimée sur les points futurs . . . 115
5.6.4 Qualité de l'ajustement et qualité de la prédiction . . . 118
5.6.5 Conclusion de l'étude comparative sur les données réelles . . . 121

6 Utilisation de méthodes d'apprentissage 125
6.1 Paramètres des méthodes d'apprentissage . . . 126
6.1.1 Méthodes d'apprentissage . . . 127
6.1.2 Variables explicatives . . . 128
6.2 Apprendre de la masse . . . 129
6.2.1 Mise en œuvre . . . 131
6.2.2 Résultats . . . 131
6.2.3 Conclusion . . . 134
6.3 Apprendre une commande de poussée . . . 138
6.3.1 Lien entre la masse et la poussée . . . 138
6.3.2 Évaluation de la qualité d'une commande de poussée . . . 140
6.3.3 Formulation du problème d'optimisation . . . 141
6.3.4 Résolution du problème d'optimisation . . . 141
6.3.5 Résultats . . . 142
6.4 Apprendre le profil de vitesse $(cas, Mach)$ . . . 149
6.4.1 Ajustement du couple $(cas, Mach)$ . . . 149
6.4.2 Apprentissage direct des cas et $Mach$ ajustés . . . 153
6.5 Comparaison des différentes méthodes . . . 161
6.5.1 Apprendre directement l'altitude . . . 161
6.5.2 Prédiction avec le modèle physique BADA . . . 165
6.5.3 Discussion sur les biais possibles . . . 167
6.6 Conclusion . . . 168

7 Généralisation à différents types avions et différentes altitudes initiales 169
7.1 Des altitudes initiales différentes . . . 169
7.1.1 Construction du jeu d'exemples . . . 169
7.2 Résultats . . . 171
7.2.1 Étude sur les exemples construits . . . 171
7.2.2 Étude sur des segments de montée plus petits . . . 181
7.3 Conclusion . . . 184

Conclusion et perspectives 185

Appendices **189**

A Preuve d'existence et de régularité de la fonction $m^*(C_1, \ldots, C_n)$ **191**
A.1 Existence de la fonction $m^*(C_1, \ldots, C_n)$ . . . 192
A.1.1 Existence d'un minimum global de $\mathcal{E}((C_1, \ldots, C_n), .)$ . . . 192
A.1.2 Unicité du minimum global . . . 193
A.2 Régularité de la fonction $m^*(C_1, \ldots, C_n)$ . . . 193
A.2.1 Théorème des fonctions implicites . . . 193
A.2.2 Application du théorème . . . 194

B Ajustement du profil (CAS,$Mach$) **197**
B.1 Le problème d'optimisation associé . . . 198
B.1.1 Preuve de convexité de $\psi_{(p,T,TAS)}$ sur I_{cas} . . . 199
B.2 Un découpage en sous-domaines . . . 200
B.2.1 Résolution sur O_M . . . 200
B.2.2 Résolution sur O_{cas} . . . 201
B.2.3 Résolution sur O_k . . . 201
B.2.4 Résolution sur F_k . . . 202
B.2.5 Conclusion sur le domaine Ω . . . 202

Glossaire **203**

Bibliographie **204**

Table des figures

1 Utiliser BADA avec les paramètres de référence. 3
2 Estimer directement la masse à partir de la trajectoire passée et en utilisant le modèle physique. 4
3 Prédire les paramètres requis pour l'utilisation du modèle physique. 5
4 Prédire directement les futures positions avec un modèle statistique appris à partir d'exemples. 5

1.1 Vue de dessus et en coupe des différentes classes d'espaces. 9
1.2 Norme de séparation pour la phase en-route. 10
1.3 Structure classique d'un prédicteur 15
1.4 Modélisation des incertitudes sur la trajectoire future 18

2.1 Angles d'Euler attachés à l'avion 23
2.2 Vitesses et angles dans un repère inertielle et dans l'air 23
2.3 Triangle des vents dans le plan horizontal 24
2.4 Profil de vitesse BADA pour un A320 30

3.1 Schéma illustrant le sur-apprentissage 39
3.2 *Forward selection* et *backward elemination* 45
3.3 Analyse en composantes principales sur un exemple 47
3.4 Représentation d'un réseau de neurones 53

4.1 Extrait d'un fichier de plans de vols COURAGE réalisé. 61
4.2 Extrait d'un fichier de plots radar Mode-C. 62
4.3 Quantification des grandeurs Mode-C. 63
4.4 Calcul naïf d'une variation temporelle. 64
4.5 Un exemple de plots manquants dans les trajectoires Mode-S. 65
4.6 Estimation de l'accélération par validation croisée 68
4.7 Estimation de la vitesse suivant l'ordre de la spline 69
4.8 Une trajectoire exemple de notre jeu d'exemples 73
4.9 Densité des H_{p0} possibles avec les données Mode-S 73
4.10 Température en fonction de l'altitude pour les données Mode-C 75
4.11 Température en fonction de l'altitude pour les données mode S 76

4.12 Vitesse Calibrated AirSpeed (CAS) en fonction de l'altitude H_p pour les données Mode-C . . . 78
4.13 Vitesse CAS en fonction de l'altitude H_p pour les données Mode-S . . . 79
4.14 Couples $(CAS_{cible}, Mach_{cible})$ ajustés pour les données Mode-C . . . 80
4.15 Couples $(CAS_{cible}, Mach_{cible})$ ajustés pour les données Mode-S . . . 81
4.16 Accélération en fonction de la vitesse pour les données Mode-C . . . 82
4.17 Accélération en fonction de la vitesse pour les données Mode-S . . . 83
4.18 Specific Energy Rate (SER) en fonction de l'altitude H_p et la vitesse CAS pour les trajectoires Mode-C . . . 85
4.19 Specific Excess Power (SEP) en fonction de l'altitude H_p et la vitesse CAS pour les trajectoires Mode-S . . . 86
4.20 Energy Share Factor (ESF) en fonction de l'altitude H_p pour les trajectoires Mode-C . . . 88
4.21 ESF en fonction de l'altitude H_p pour les trajectoires Mode-S . . . 89

5.1 Estimer la masse à partir de BADA et de la trajectoire passée. . . . 93
5.2 Déroulement de la recherche dichotomique . . . 94
5.3 Estimation point par point de la masse . . . 98
5.4 Algorithme de Newton par intervalle . . . 102
5.5 Sensibilité des méthodes d'estimation de la masse aux erreurs sur la température T . . . 104
5.6 Sensibilité des méthodes d'estimation de la masse aux erreurs sur l'altitude H_p . . . 105
5.7 Sensibilité des méthodes d'estimation de la masse aux erreurs sur la vitesse V_a . . . 105
5.8 Sensibilité des méthodes d'estimation de la masse aux erreurs sur l'accélération $\frac{dV_a}{dt}$ . . . 106
5.9 Sensibilité des méthodes d'estimation de la masse aux erreurs sur le taux de montée $\frac{dH_p}{dt}$ . . . 106
5.10 Erreurs commises sur les masses avec un bruit sur la température T . . . 107
5.11 Poussée obtenue en fonction de ΔT . . . 108
5.12 Cas d'étude pour l'estimation de la masse . . . 110
5.13 Différence entre la puissance calculée et observée sur un exemple . . . 113
5.14 Différence entre la puissance calculée et observée sur un autre exemple . . . 114
5.15 Distribution de la masse estimée sur les points passés des trajectoires Mode-C 115
5.16 Distribution de la masse estimée sur les points passés des trajectoires Mode-S 116
5.17 Masse estimée sur les points futurs en fonction de la distance à parcourir . . . 117
5.18 Différence entre les masses estimée sur les points futurs et passés pour les trajectoires Mode-C . . . 119
5.19 Différence entre les masses estimée sur les points futurs et passés pour les trajectoires Mode-S . . . 119
5.20 Erreur sur l'altitude en fonction de l'erreur sur la puissance pour les trajectoires Mode-C . . . 120

5.21 Erreur sur l'altitude en fonction de l'erreur sur la puissance pour les trajectoires Mode-S . . . 121
5.22 Erreur sur la puissance sur les points futurs en fonction de l'erreur sur les points passés pour les trajectoires Mode-C . . . 122
5.23 Erreur sur la puissance sur les points futurs en fonction de l'erreur sur les points passés pour les trajectoires Mode-S . . . 123

6.1 Prédire les paramètres manquants pour améliorer la prévision . . . 126
6.2 Erreur de puissance calculée sur les points futurs avec la masse prédite et la masse optimale . . . 136
6.3 Exemple illustrant la puissance calculée avec différentes masses . . . 137
6.4 La puissance calculée en fonction de la masse et de la commande de poussée 139
6.5 Deux commandes de poussée expliquant parfaitement la puissance observée 140
6.6 Commandes de poussée apprises . . . 143
6.7 Commandes de poussée apprises . . . 144
6.8 Commandes de poussée apprises suivant la variante de l'appareil . . . 144
6.9 Commandes de poussée apprises suivant la variante de l'appareil . . . 145
6.10 Qualité de l'ajustement d'un profil $(cas, Mach)$ pour les trajectoires Mode-C 151
6.11 Qualité de l'ajustement d'un profil $(cas, Mach)$ pour les trajectoires Mode-S 152
6.12 Erreur obtenue avec le couple prédit $(cas, Mach)$ en fonction de l'erreur optimale. . . . 160
6.13 Prédire en apprenant directement l'altitude finale . . . 162

7.1 Construction d'exemples à partir d'une trajectoire . . . 170
7.2 Fonctions de répartitions de la valeur absolue de l'erreur obtenue avec $BADA_{ref}$ ou $BADA_{\text{GBM}}$ pour différents types avions . . . 178
7.3 Valeur absolue de l'erreur en fonction de l'altitude initiale de la prédiction 179
7.4 Erreur en fonction du temps pour différents types avions . . . 180

Liste des tableaux

4.1 Écarts entre les trajectoires lissées et les plots observés pour les trajectoires Mode-C . . . 70
4.2 Écarts entre les trajectoires lissées et les plots observés pour les trajectoires Mode-S . . . 71

5.1 Distribution des paramètres utilisés pour générer les trajectoires simulées . 103
5.2 Sachant $V_a^{(obs)}$ et la masse estimée, erreurs sur l'altitude pour les trajectoires Mode-C . . . 111
5.3 Sachant $V_a^{(obs)}$ et la masse estimée, erreurs sur l'altitude pour les trajectoires Mode-S . . . 112
5.4 Écarts d'altitude en utilisant la masse estimée sur les points futurs 116
5.5 Écarts d'altitude en utilisant la masse estimée sur les points futurs 118
5.6 Écarts entre la masse « réelle »et celle estimée pour les trajectoires Mode-C 118
5.7 Écarts entre la masse « réelle »et celle estimée pour les trajectoires Mode-S 118

6.1 Méthodes d'apprentissage utilisés et grille d'hyper-paramètres associée . . 127
6.2 Ce tableau décrit les variables utilisées par les méthodes d'apprentissage. . 129
6.3 Ce tableau résume les différents jeux de variables utilisés. 130
6.4 Écarts entre la masse « réelle »et celle prédite pour les trajectoires Mode-C 132
6.5 Écarts entre la masse « réelle »et celle prédite pour les trajectoires Mode-S 133
6.6 Sachant $V_a^{(obs)}$ et la masse prédite, erreurs sur l'altitude pour les trajectoires Mode-C . . . 134
6.7 Sachant $V_a^{(obs)}$ et la masse prédite, erreurs sur l'altitude pour les trajectoires Mode-S . . . 135
6.8 Sachant $V_a^{(obs)}$, erreurs sur l'altitude prédite avec une commande de poussée apprise pour les trajectoires Mode-C . . . 146
6.9 Sachant $V_a^{(obs)}$, erreurs sur l'altitude prédite avec une commande de poussée apprise pour les trajectoires Mode-S . . . 147
6.10 Statistiques, en W/kg, sur l'erreur $e^{[11:51]}(\hat{m}_{11,futur})$. Différentes commandes de poussée sont testées. 148
6.11 Statistiques, en kg, sur l'erreur $\hat{m}_{11} - \hat{m}_{11,futur}$. Différentes commandes de poussée sont testées. La masse $\hat{m}_{11}$ est estimée par la méthode des moindres carrés. 148

6.12 Statistiques en nœuds pour les trajectoires Mode-C sur la différence entre la *cas* ajustée sur les points futurs et la *cas* choisie à partir des informations disponibles au moment de la prédiction. 154
6.13 Statistiques en nœuds pour les trajectoires Mode-S sur la différence entre la *cas* ajustée sur les points futurs et la *cas* choisie à partir des informations disponibles au moment de la prédiction. 155
6.14 Statistiques pour les trajectoires Mode-C, sans dimensions, sur la différence entre le $Mach$ ajusté sur les points futurs et le $Mach$ choisi à partir des informations disponibles au moment de la prédiction. 156
6.15 Statistiques pour les trajectoires Mode-S, sans dimensions, sur la différence entre le $Mach$ ajusté sur les points futurs et le $Mach$ choisi à partir des informations disponibles au moment de la prédiction. 157
6.16 Statistiques en nœuds sur $(V_{acible} - V_a)\,(t \geqslant 0)$ pour les trajectoires Mode-C. 158
6.17 Statistiques en nœuds sur $(V_{acible} - V_a)\,(t \geqslant 0)$ pour les trajectoires Mode-S. 159
6.18 Écarts entre l'altitude« réelle »et celle prédite pour les trajectoires Mode-C 163
6.19 Écarts entre l'altitude« réelle »et celle prédite pour les trajectoires Mode-S 164
6.20 Statistiques, en pieds, sur l'écart entre les trajectoires prédites et les plots observés pour les trajectoires Mode-C. Ces statistiques sont calculées sur l'ensemble des valeurs $\left(H_p^{(pred)}(\hat{m}_{11}) - H_p^{(obs)}\right)\,(t = 600 \text{ s})$. Chacune de ces valeurs est associée à une trajectoire. Différentes commandes de poussée sont testées. 165
6.21 Statistiques, en pieds, sur l'écart entre les trajectoires prédites et les plots observés pour les trajectoires Mode-S. Ces statistiques sont calculées sur l'ensemble des valeurs $\left(H_p^{(pred)}(\hat{m}_{11}) - H_p^{(obs)}\right)\,(t = 600 \text{ s})$. Chacune de ces valeurs est associée à une trajectoire. Différentes commandes de poussée sont testées. 166

7.1 Statistiques sur l'erreur faite en prédisant à partir de différentes altitudes initiales pour les A319 172
7.2 Statistiques sur l'erreur faite en prédisant à partir de différentes altitudes initiales pour les A320 172
7.3 Statistiques sur l'erreur faite en prédisant à partir de différentes altitudes initiales pour les A321 173
7.4 Statistiques sur l'erreur faite en prédisant à partir de différentes altitudes initiales pour les A332 173
7.5 Statistiques sur l'erreur faite en prédisant à partir de différentes altitudes initiales pour les B737 174
7.6 Statistiques sur l'erreur faite en prédisant à partir de différentes altitudes initiales pour les B744 174
7.7 Statistiques sur l'erreur faite en prédisant à partir de différentes altitudes initiales pour les B772 175
7.8 Statistiques sur l'erreur faite en prédisant à partir de différentes altitudes initiales pour les E145 175

7.9 Statistiques sur l'erreur faite en prédisant à partir de différentes altitudes initiales pour les F100 176
7.10 Statistiques sur le nombre de segments de montée 181
7.11 Statistiques sur l'erreur faite en prédisant à 5 min à partir de différentes altitudes initiales 183

Liste des algorithmes

6.1 Algorithmes de sélection d'hyper-paramètres 128

Introduction

Problématique

Les prévisions de trafic en Europe effectuées par Eurocontrol [STATFOR 13] indiquent une tendance à la hausse malgré la crise économique. Tous les scénarios présentés dans le document [STATFOR 13] prévoient une augmentation du nombre de vols d'au moins 20 % d'ici 2035. Deux gros projets en cours dans le monde de l'Air Traffic Management (ATM), le projet Single European Sky ATM Research (SESAR) ([SESAR Consortium 07]) et le projet NextGen ([Swenson 06]), visent à une augmentation de la capacité et de la sécurité tout en diminuant le coût du contrôle aérien. Cette amélioration passe par l'utilisation d'outils d'aide à la décision. Certains de ces outils reposent sur une prédiction de trajectoire de qualité.

Parmi ces outils, certains visent à aider le contrôleur aérien dans la détection et la résolution de conflits. C'est la tâche principale du contrôleur aérien dans l'espace aérien supérieur [2], elle consiste à optimiser l'écoulement du trafic aérien tout en empêchant les collisions entre aéronefs. Cet anti-abordage est assuré en garantissant un espacement minimum d'au moins 5 NM horizontalement ou d'au moins 1 000 ft verticalement. Des algorithmes de détection et résolution de conflits ont été développés par les laboratoires sur le site de l'École Nationale de l'Aviation Civile (ÉNAC) ([Durand 96, Granger 02]). Ces algorithmes reposent sur l'évaluation d'un grand nombre de scénarios différents. Parmi ces scénarios de manœuvres envisagés, cet algorithme va chercher celui qui garantit des trajectoires futures séparées tout en minimisant la somme des coûts associés aux manœuvres choisies. Le test de ces différents scénarios va donc nécessiter de connaître les trajectoires futures. Plus l'incertitude est grande sur ces trajectoires futures et plus le choix des manœuvres garantissant la séparation est restreint. Cette restriction se traduit par une augmentation du coût du scénario optimal. Ainsi, comme cet algorithme est mis en œuvre dans les systèmes sols, il est important de pouvoir fournir une prédiction de qualité en utilisant uniquement les informations disponibles au sol.

Les projets NextGen et SESAR s'appuient, entre autres, sur le concept d'une trajectoire 4D négociée entre le pilote et le contrôleur. Toutefois, la mise en œuvre de ce concept de trajectoires 4D nécessite des investissements pour que cet échange entre l'avion en vol et le contrôleur au sol puisse avoir lieu. De même, les compagnies devront investir dans des

2. Espace aérien du niveau 195 (exclu) au niveau 660.

Flight Management System (FMS) capables de suivre de telles trajectoires 4D.

En France, dans le système actuel et ce depuis les années soixante, la communication entre pilote et contrôleur se fait par un système de radiocommunication. Cette thèse explore les moyens d'améliorer, avec les systèmes actuellement opérationnels, la qualité de la prédiction de trajectoire à un horizon de 10 min, horizon pertinent pour la détection et la résolution de conflits. On va en particulier s'intéresser à la prédiction de la phase de montée pour laquelle les performances varient grandement entre avions de type différents et même entre avions de même type.

Quelques travaux ([Le Fablec 99, Tastambekov 14, Ghasemi Hamed 14]) ont déjà étudiés la prédiction de trajectoire à l'aide de méthodes purement statistiques en utilisant un jeu de plusieurs trajectoires. D'autres travaux ([Lymperopoulos 06, Schultz 12]) utilisent uniquement les points passés de la trajectoire considérée pour trouver les paramètres nécessaires à l'usage d'un modèle physique. Notre travail se situe entre ces deux approches, on exploite un jeu de plusieurs trajectoires pour trouver les paramètres à utiliser dans un modèle physique.

Approche adoptée dans cette thèse

L'utilisation d'un modèle physique requiert la connaissance de la masse, de la commande de poussée et du profil de vitesse suivi. Ces trois paramètres sont inconnus des systèmes sols. Le projet BADA d'Eurocontrol répond partiellement à ce besoin. Il fournit un modèle physique de l'avion et des valeurs de référence pour les paramètres inconnus. La figure 1 résume le processus permettant le calcul des altitudes futures à partir des valeurs de référence BADA. Cette méthode très largement utilisée dans les systèmes actuels, est imprécise du fait de notre méconnaissance des valeurs réelles des paramètres du modèle physique. Ces données pourraient aisément être transmise au sol par les avions mais elles sont jugées trop « sensibles » par la majorité des compagnies aériennes. L'idée centrale que nous allons développer dans cette thèse consiste à estimer plus finement ces paramètres inconnus, soit directement à partir des points passés de la trajectoire et en utilisant le modèle physique de l'avion, soit en apprenant, à partir d'une base d'exemples, des modèles permettant de prédire les valeurs de ces paramètres. Dans ce dernier cas, nous utiliserons des techniques d'apprentissage supervisé.

Les points passés de la trajectoire peuvent être exploités pour inférer une masse. Un avion ayant une vitesse ascensionnelle élevée par rapport à ce qui est modélisé par le modèle physique a probablement une masse faible. La figure 2 résume ce processus permettant l'estimation de la masse à partir des points passés et d'un modèle physique. Cette idée, abordée par [Schultz 12], est également développée dans cette thèse.

Dans l'approche précédente, la distance à parcourir est complètement ignorée alors qu'elle est fortement corrélée à la quantité de carburant nécessaire et donc à la masse de l'appareil. En effet, uniquement les variables intervenant dans le modèle physique sont exploitées. Pour exploiter toutes les variables disponibles on va devoir utiliser des relations qui n'apparaissent pas dans le modèle physique. Pour découvrir ces relations, on utilise

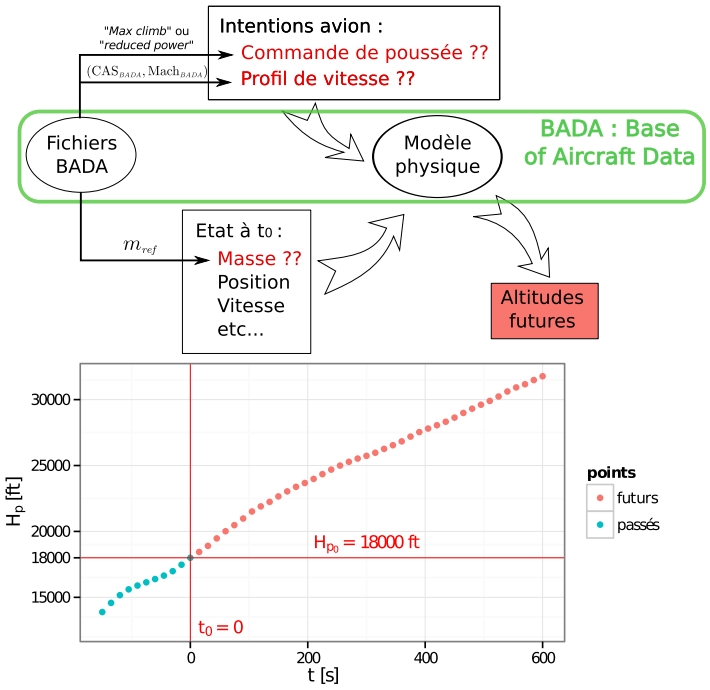


FIGURE 1 – Utiliser BADA avec les paramètres de référence.

des méthodes d'apprentissage artificiel qui vont apprendre ces relations à partir d'une base d'exemples. De cette façon, nous obtiendrons des modèles permettant de prédire la masse de l'avion, sa commande de poussée et son profil de vitesse. Cette approche est celle adoptée dans notre thèse. La figure 3 résume ce processus permettant la prédiction des paramètres requis à partir des points passés, d'une base d'exemples et d'un modèle physique.

Les résultats de notre approche sont comparés à ceux obtenus d'une part avec la méthode de référence BADA et d'autre part avec une approche purement statistique où l'on apprend directement l'altitude future de l'avion au lieu des paramètres du modèle physique. Dans cette approche, déjà explorée dans d'autres travaux ([Le Fablec 99, Tastambekov 14, Ghasemi Hamed 14]), les relations apprises à l'aide de méthodes d'apprentissage décrivent directement la cinématique de l'avion ; elles décrivent les positions futures et non les paramètres nécessaire à l'utilisation d'un modèle physique. La figure 4 résume cette approche.

Les travaux effectués dans le cadre de cette thèse ont fait l'objet de plusieurs publications et communications ([Alligier 11, Alligier 12, Alligier 13b, Alligier 13a, Alligier 14]).

Organisation du document

Cette thèse est constituée de sept chapitres. Le chapitre 1 décrit la gestion du trafic aérien, la prévision de trajectoires et ses enjeux. Le chapitre 2 présente la modélisation physique de la prédiction de trajectoire. Le modèle physique BADA y est détaillé. Le

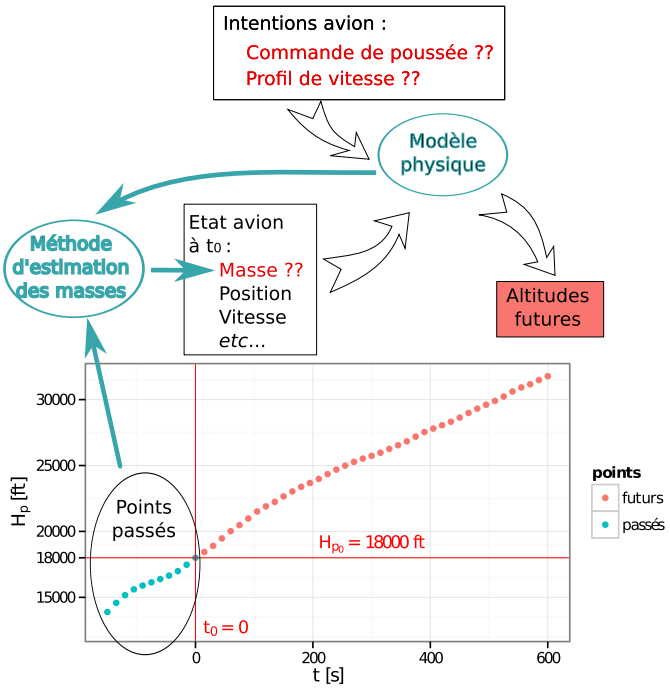


FIGURE 2 – Estimer directement la masse à partir de la trajectoire passée et en utilisant le modèle physique.

chapitre 3 présente les principes associés à l'apprentissage supervisé ainsi que différentes méthodes dédiées à cette tâche. Le chapitre 4 décrit les données utilisées dans cette étude, en particulier le filtrage et lissage des données. Une analyse des trajectoires finalement obtenues est faite. Le chapitre 5 décrit les méthodes d'estimation de la masse développées dans cette thèse. Les masses sont estimées uniquement à partir des points passés et du modèle physique BADA. Le chapitre 6 décrit la mise en œuvre des méthodes d'apprentissage pour prédire les paramètres inconnus et nécessaires à l'utilisation du modèle physique BADA. Dans ce chapitre, on compare les résultats obtenus entre la prédiction BADA renseigné avec nos méthodes et la prédiction directe de l'altitude à un horizon de 10 min. Le chapitre 7 élargit l'étude à différents types avions et différentes altitudes initiales de prédiction. Il est aussi vérifié que les modèles obtenus pour une prédiction à un horizon de 10 min sont également pertinents pour prédire sur un horizon plus faible.

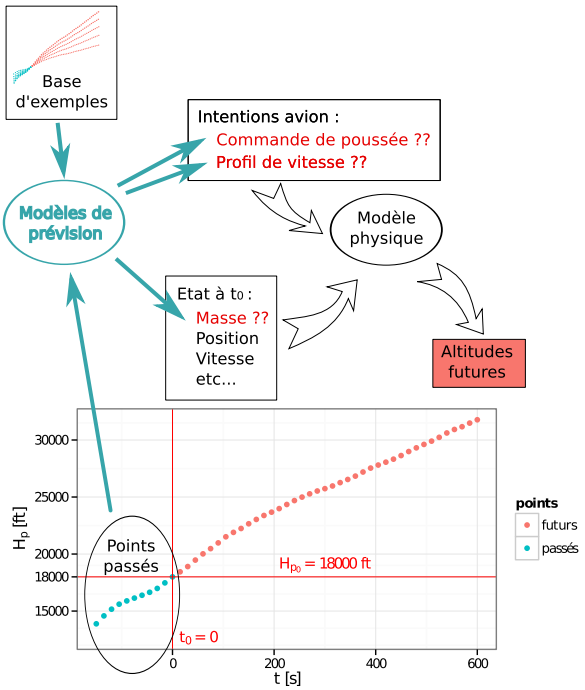


FIGURE 3 – Prédire les paramètres requis pour l'utilisation du modèle physique.

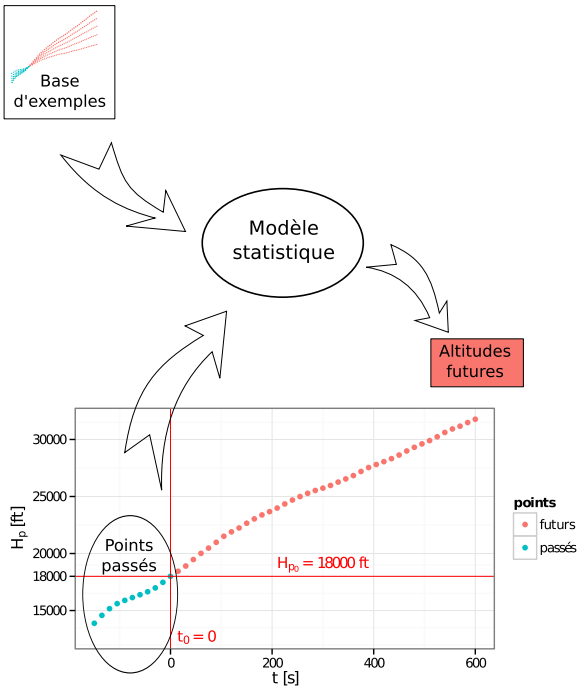


FIGURE 4 – Prédire directement les futures positions avec un modèle statistique appris à partir d'exemples.

Chapitre 1

Contexte

Sommaire

1.1 Gestion du trafic aérien 7
1.1.1 Différents types de vols 8
1.1.2 Différentes classes d'espace aérien 8
1.2 Contrôle du trafic aérien 10
1.2.1 Vocabulaire et unités de mesure 10
1.2.2 Méthodes du contrôle 11
1.3 Évolutions dans la gestion du trafic aérien 12
1.3.1 Évolution du trafic aérien en Europe 12
1.3.2 Évolution des concepts de gestion du trafic aérien 12
1.4 Prévision de trajectoires et enjeux associés pour l'ATM 13
1.4.1 Méthodes existantes 14
1.4.2 Détection et résolution de conflits 16
1.5 Cadre de travail 17

Ce chapitre décrit les prédicteurs de trajectoires et les enjeux associés à la gestion du trafic aérien. L'augmentation du trafic aérien européen envisagée par Eurocontrol ([STATFOR 13]) pousse au développement de nouveaux concepts et outils informatiques s'appuyant sur la prédiction de trajectoire. C'est dans ce cadre que s'inscrit ce travail de thèse.

1.1 Gestion du trafic aérien

Les services de la circulation aérienne se composent de trois services permettant d'assurer l'exécution sûre et efficace des vols [SCA 06] :

1. Le service du contrôle de la circulation aérienne qui a pour objet :
   – d'empêcher les collisions entre aéronefs en vol ;
   – d'empêcher les collisions entre aéronefs sur l'aire de manœuvre et les obstacles se trouvant sur cette aire ;

– d'accélérer et ordonner la circulation aérienne ;

2. Le service d'information de vol qui fournit les avis et les renseignements utiles à l'exécution sûre et efficace des vols ;
3. Le service d'alerte qui prévient et renseigne les organismes appropriés lorsque des aéronefs ont besoin de l'aide des organismes de recherches et de sauvetage.

L'ensemble des services du contrôle de la circulation n'est pas assuré pour tous les espaces aériens ni tous les types de vols.

1.1.1 Différents types de vols

Dans l'espace aérien, on distingue deux types de vols : le vol Visual Flight Rules (VFR) autrement appelé *vol à vue* et le vol Instrument Flight Rules (IFR) autrement appelé *vol aux instruments*. Le *vol à vue* nécessite des conditions de visibilité et de distance par rapport aux nuages. Un tel vol est donc très impacté par les conditions atmosphériques. Ceci est inhérent à la règle du « voir et éviter » régissant l'anti-abordage faute d'une séparation assurée par le service de contrôle. Ce type de vol est adapté au vol de tourisme. Le *vol aux instruments* est un vol pour lequel la navigation est faite à l'aide des instruments de radionavigation à bord de l'appareil. La séparation est assurée par les services de contrôle. Ce type de vol est moins dépendant des conditions atmosphériques, il est plus adapté aux vols commerciaux.

1.1.2 Différentes classes d'espace aérien

L'espace aérien est découpé en zones de différentes classes. Ces zones sont des volumes. Il existe 7 classes d'espaces aériens nommés de A à G. Suivant la classe de l'espace et le type de vol, le service de contrôle rendu est différent. Par exemple, une séparation entre tous les vols est assurée par le service de contrôle dans un espace de classe A ou B. Pour un espace de classe C, seul l'espacement entre avions VFR n'est pas assuré. Pour un espace de classe D ou E, seul l'espacement entre avions IFR est assuré. Pour les espaces de classe F et G, le service de contrôle n'est pas assuré. Le service assuré dans l'espace aérien, i.e. la classification de cet espace, est décidé en considérant le type de trafic en cause, la densité de circulation aérienne, les conditions atmosphériques et toutes autres conditions particulières.

En dessous du FL115 [1], l'espace aérien est de classe G sauf à proximité d'aéroports qui sont protégés par des zones Control Traffic Region (CTR) et Terminal Manœuvring Area (TMA) dans lesquelles le service de contrôle est assuré. Les plus gros aéroports sont entourés par des espaces de classe A ; c'est le cas par exemple des aéroports de la région parisienne comme illustré par la figure 1.1. Au-dessus du FL115 et jusqu'au FL195, l'espace aérien est principalement de classe D. Au-dessus du FL195 et jusqu'au FL660, il y a une unique zone de classe C.

1. L'altitude en aéronautique est mesurée en Flight Level (FL) ; elle correspond à l'altitude pression en centaines de pieds dans l'atmosphère International Standard Atmosphere (ISA).

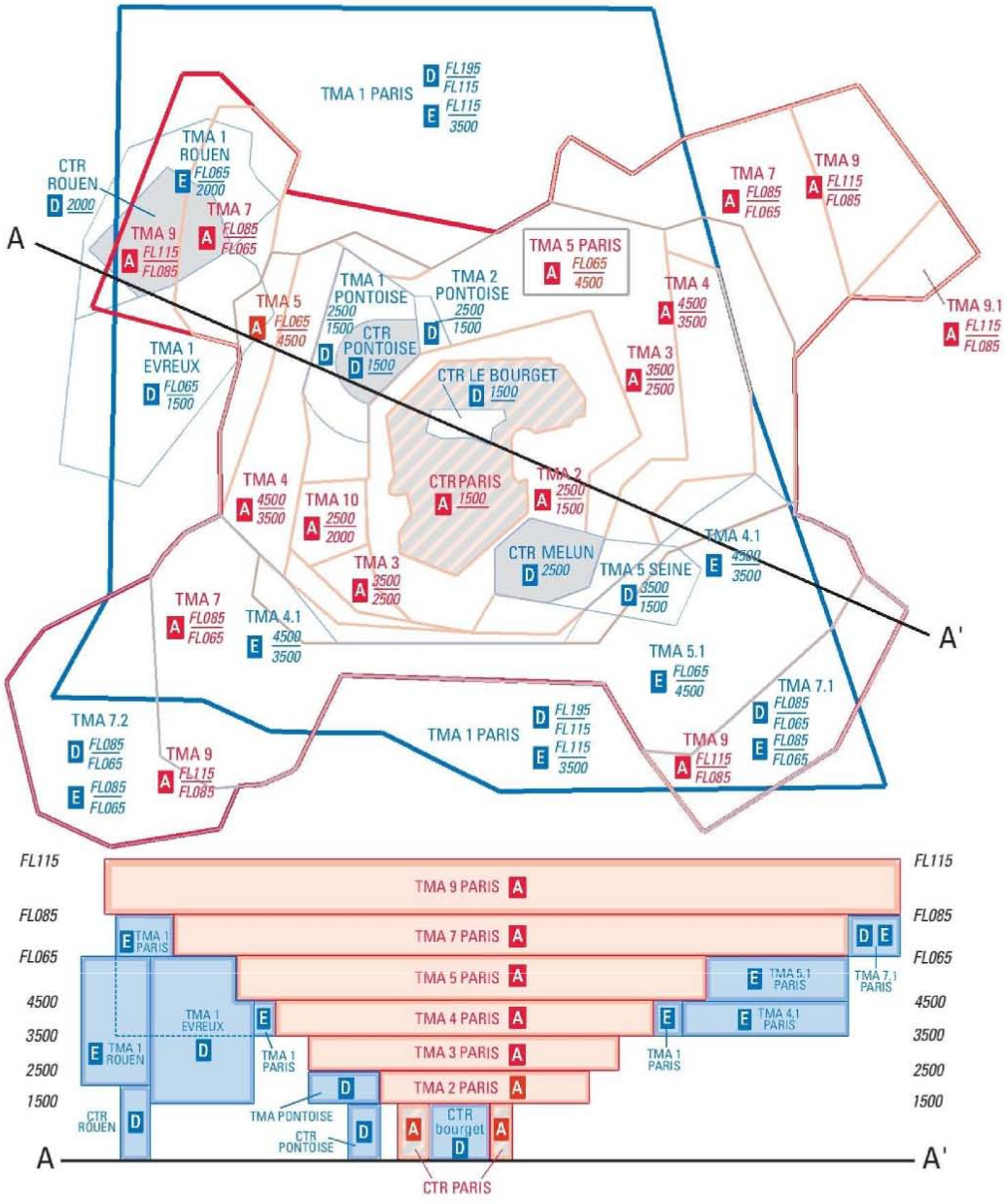


FIGURE 1.1 – Vue de dessus et en coupe des différentes classes d'espaces aérien de la région parisienne.

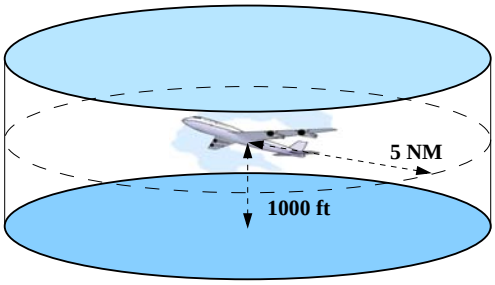


FIGURE 1.2 – Norme de séparation pour la phase en-route. Aucun autre avion ne doit se trouver dans ce volume.

1.2 Contrôle du trafic aérien

Le but premier du service de contrôle est d'empêcher les collisions entre aéronefs. Le service de contrôle vise aussi à optimiser l'écoulement du trafic sans remettre en cause ce but premier. Pour pouvoir empêcher les collisions entre aéronefs, le service de contrôle vise à maintenir une séparation géométrique entre avions illustrée par la figure 1.2. Cette norme de séparation impose que deux avions soient toujours espacés d'au moins 5 NM horizontalement ou d'au moins 1 000 ft verticalement.

1.2.1 Vocabulaire et unités de mesure

Les unités de mesure utilisées en aéronautique ne sont pas des unités du système international. Les longueurs ne sont pas mesurées en mètres. Par exemple, la distance entre deux points du globe est exprimée en Nautical Mile (NM).

Les altitudes sont exprimées en pieds (ft) ou en Flight Level (FL), avec 1 FL qui équivaut à 100 ft. Les FL sont uniquement utilisés pour la grandeur *altitude pression géopotentielle*. Cette grandeur est déterminée à l'aide la pression statique p mesurée par l'avion. L'*altitude pression géopotentielle* est l'altitude géopotentielle pour laquelle la pression modélisée dans le modèle d'atmosphère ISA est égale à p. Cette *altitude pression géopotentielle* est la grandeur servant à la définition de la norme de séparation verticale. De même, les pilotes évoluant à une altitude suffisamment élevée utilisent cette grandeur pour se situer verticalement dans l'atmosphère. Les différentes grandeurs associées à une notion d'altitude sont détaillées dans la sous-section 2.1.

Concernant les vitesses, elles sont mesurées en *nœuds* (kts), avec 1 kts qui équivaut à 1 NM/h. Il y a au moins trois grandeurs différentes quantifiant la vitesse de l'avion dans l'air. La TAS (True Airspeed) est la vitesse relative de l'avion dans la masse d'air. La CAS (Calibrated Airspeed) est la TAS nécessaire au niveau du sol pour avoir la même pression dynamique que celle mesurée par l'avion. Si l'on écarte les erreurs d'instruments, la CAS est la vitesse indiquée sur les instruments de bord du pilote. Le nombre de Mach est le ratio entre la TAS et la vitesse du son dans l'air. Ces deux dernières vitesses sont celles utilisées par le pilote pour mettre en œuvre son appareil.

1.2.2 Méthodes du contrôle

Les espaces aériens contrôlés sont divisés en secteurs. Cette division en secteur varie au cours de la journée en fonction du trafic. Le service de contrôle sur chaque secteur est assuré par une paire de contrôleurs aériens. Le contrôleur dit *organique* planifie le travail dans les minutes à venir. En effet, chaque avion qui va entrer dans son secteur donne lieu à l'impression d'une bandelette papier appelée *strip* sur laquelle figure les informations utiles et disponibles concernant ce nouvel entrant. Le contrôleur organique réceptionne ce *strip* et s'assure que l'entrée de l'avion dans le secteur puisse se faire dans de bonnes conditions. Si ce n'est pas le cas, il peut demander au secteur d'où vient l'avion des modifications de trajectoires. De manière symétrique, il reçoit les demandes venant des secteurs recevant les avions sortant de son secteur. En résumé, il coordonne les entrées et sorties de son secteur.

Il travaille avec le contrôleur *radar* qui s'assure du respect des normes de séparation et de l'optimisation des flux d'avions de son secteur. Pour ce faire, il utilise une visualisation radar présentant une vue de dessus de l'état du trafic aérien dans son secteur. La position de chaque avion est visualisée par un plot radar sur lequel il peut lire l'altitude ainsi que l'évolution de celle-ci. Si le contrôleur anticipe une perte de séparation à venir entre deux avions, i.e. un *conflit*, il ordonne à au moins un des avions d'effectuer une manœuvre comme par exemple un changement de cap, un changement de vitesse, une anticipation de descente ou une interruption de montée. Le contrôleur peut anticiper des conflits une vingtaine de minutes en amont ; la demande de manœuvre résolvant le conflit peut être faite quelques minutes avant le conflit.

Le trafic aérien actuel est organisé en réseaux de routes aériennes. Ces routes empruntées par les avions sont constituées de segments convergents vers des points. Des équipements radio-électriques peuvent être placés au sol en ces points permettant ainsi à l'avion de se situer à l'aide d'équipements de radionavigation à bord. Cette organisation facilite la tâche du contrôleur car les conflits dûs à des trajectoires convergentes ne sont localisés qu'aux croisements de ces routes.

Le travail effectué peut toutefois être très différent suivant la nature du secteur et ce, malgré une définition commune du rôle du contrôleur aérien. En effet, on distingue trois types de contrôle :

- le contrôle en route : il gère la progression des avions en dehors des zones proches des aéroports. Il contrôle en particulier le trafic le long des routes aériennes.
- le contrôle d'approche : il gère la phase de descente de l'avion jusqu'à 6 ou 10 Nm de la piste. Son but est le séquencement et l'alignement des avions sur la piste.
- le contrôle d'aérodrome : il gère les avions alignés par le contrôle d'approche. En particulier il gère les pistes au décollage comme à l'atterrissage.

1.3 Évolutions dans la gestion du trafic aérien

1.3.1 Évolution du trafic aérien en Europe

Les prévisions de trafic en Europe effectuées par Eurocontrol [STATFOR 13] indiquent une tendance à la hausse malgré la crise économique. Plusieurs scénarios sont étudiés à l'horizon 2035. Ils prennent en compte différents paramètres impactant l'offre et la demande comme la démographie, l'infrastructure ferroviaire, les conditions économiques et les stratégies des compagnies aériennes. Parmi les quatre scénarios envisagés, le plus pessimiste conduit à une augmentation du nombre de vols de 20 %. Le scénario le plus probable selon ce document conduit à une augmentation de 50 % du nombre de vols de 2012 à 2035. Ce document prédit que la capacité des aéroports sera un des facteurs pouvant empêcher de répondre à la demande en 2035.

Cette augmentation du nombre de vols conduit *a priori* à une augmentation du nombre de vols dans une même zone au même moment. Les contrôleurs d'un secteur peuvent toutefois ne gérer qu'un nombre limité d'avions. Découper la zone en plusieurs secteurs n'est pas toujours une solution. En effet, plus les secteurs sont petits, moins il y a d'espace pour que les manœuvres puissent avoir lieu à l'intérieur du secteur et plus le travail de coordination entre secteurs devient important, limitant ainsi le nombre d'avions gérables.

1.3.2 Évolution des concepts de gestion du trafic aérien

Des concepts opérationnels pouvant résoudre ce problème ont été proposés. Dans le concept *sector-less* proposée par [Duong 01] la notion de secteur disparaît au profit d'une notion de groupe d'avions. Le contrôleur ne gère plus un secteur mais un groupe d'avions.

D'autres concepts s'affranchissent du système de contrôle *centralisé* tel qu'en vigueur aujourd'hui. Dans des concepts comme le *free flight* ([RTCA 95]) les avions assurent eux-mêmes leur séparation grâce à un équipement de bord permettant la négociation automatique de manœuvres résolvant les conflits entre avions. C'est une approche *distribuée* dans laquelle les décisions sont prises sans avoir une vision globale du trafic. De ce fait, les résolutions de conflits trouvées peuvent ne pas être « optimales » pour le trafic pris dans son ensemble.

Le concept *free-route* proposé par [dos Santos 02] permet aux avions de sortir du réseau de routes aériennes. Le contrôleur reste toutefois responsable du maintien de la séparation. Ce concept n'implique pas un changement de paradigme complet, le rôle du contrôleur reste le même. Ce concept est d'ailleurs en essai opérationnel depuis 2011 dans l'espace aérien contrôlé par Maastricht ([Eurocontrol 11]). Dès 2013, cet essai a été étendu au Functional Airspace Block Europe Central (FABEC) pour des niveaux supérieurs à FL365 ([Eurocontrol 12b]) pour des situations de trafic peu dense, la nuit et le week-end.

Les deux gros projets en cours dans le monde de l'ATM sont le projet Européen SESAR ([SESAR Consortium 07]) et son équivalent aux États-Unis *NextGen* ([Swenson 06]). Ces projets ambitieux visent à une augmentation de la capacité et de la sécurité tout en diminuant le coût du contrôle aérien. Ces projets on été lancés dans les années 2000 avec un

déploiement à l'horizon 2020 pour SESAR et 2025 pour *NextGen*. Ces projets s'appuient, entre autres, sur le concept d'une trajectoire 4D négociée entre le pilote et contrôleur aérien. Une fois la trajectoire acceptée, le pilote s'engage à se conformer à cette trajectoire. De même, le contrôleur peut être automatiquement prévenu si l'avion s'écarte de la trajectoire négociée. Le premier essai d'une implémentation partielle de ce concept a été effectué en 2012 [Mutuel 13]. Une trajectoire 3D est négociée et une contrainte de temps est fixée sur un des points de la trajectoire.

Toutefois, la mise en œuvre de ce concept de trajectoires 4D nécessite des investissements pour que cet échange entre l'avion en vol et le contrôleur au sol puisse avoir lieu. Actuellement, la communication entre pilote et contrôleur se fait par un système de communication radio et ce depuis les années soixante. De même, les compagnies devront investir dans des FMS, les ordinateurs de bord des avions, capables de suivre de telles trajectoires 4D.

1.4 Prévision de trajectoires et enjeux associés pour l'ATM

La prévision de trajectoires permet de construire une représentation des positions futures d'un avion à partir des informations disponibles. Ces positions futures sont utiles dans la gestion du trafic aérien. Elles permettent au contrôleur de prévoir des régulations si les flux d'avions prévus sont trop importants. Elles sont indispensables aux outils informatisés permettant d'aider le contrôleur dans la prise de décision et la surveillance du trafic. Ces outils, associés à une logique d'emploi adaptée, permettent au contrôleur de gérer plus de trafic.

Dans le document [Eurocontrol 10], on distingue la *planned trajectory* qui prédit le vol du décollage à l'atterrissage en s'appuyant sur les procédures et routes aériennes. Elle prévoit et tient compte des autorisations du contrôle aérien nécessaires au déroulement du vol mais n'anticipe pas d'éventuelles instructions pour résoudre un conflit par exemple. La *planned trajectory* peut être calculée bien avant que le vol ait lieu. Cela permet par exemple de déterminer quand un avion rentre dans un secteur et ainsi de calculer la charge prévue d'un secteur. Ce calcul peut être mis à jour avec l'heure de départ effective et aussi au cours du vol. Cette prédiction peut aussi être utilisée dans un outil appelé Arrival MANager (AMAN) qui va aider le contrôleur à séquencer les avions sur les pistes à l'arrivée.

La *tactical trajectory* tient compte de toutes les clairances et instructions déjà données. Elle permet de construire la trajectoire future sur un horizon de 5-10 minutes. Cette prédiction de trajectoire est associée à une enveloppe qui doit contenir la trajectoire future. Cette enveloppe peut être utilisée pour détecter les conflits. Au-delà de la simple détection, des outils à venir pourraient utiliser la prédiction de trajectoire pour suggérer au contrôleur des trajectoires alternatives permettant la résolution d'un conflit. Cette capacité suppose que l'outil de prédiction de trajectoire est capable de tester divers scénarios incluant de nouvelles instructions. L'efficacité de ces outils, et donc au final l'utilité et l'acceptabilité de

ces outils pour les contrôleurs aériens, dépend fortement de la performance du prédicteur de trajectoires.

Il est toutefois à noter que dans le système français actuel, les clairances sont données via la radio VHF et notées sur un *strip* ; le système informatique de gestion du trafic aérien n'est donc actuellement pas renseigné sur les clairances et instructions déjà données.

1.4.1 Méthodes existantes

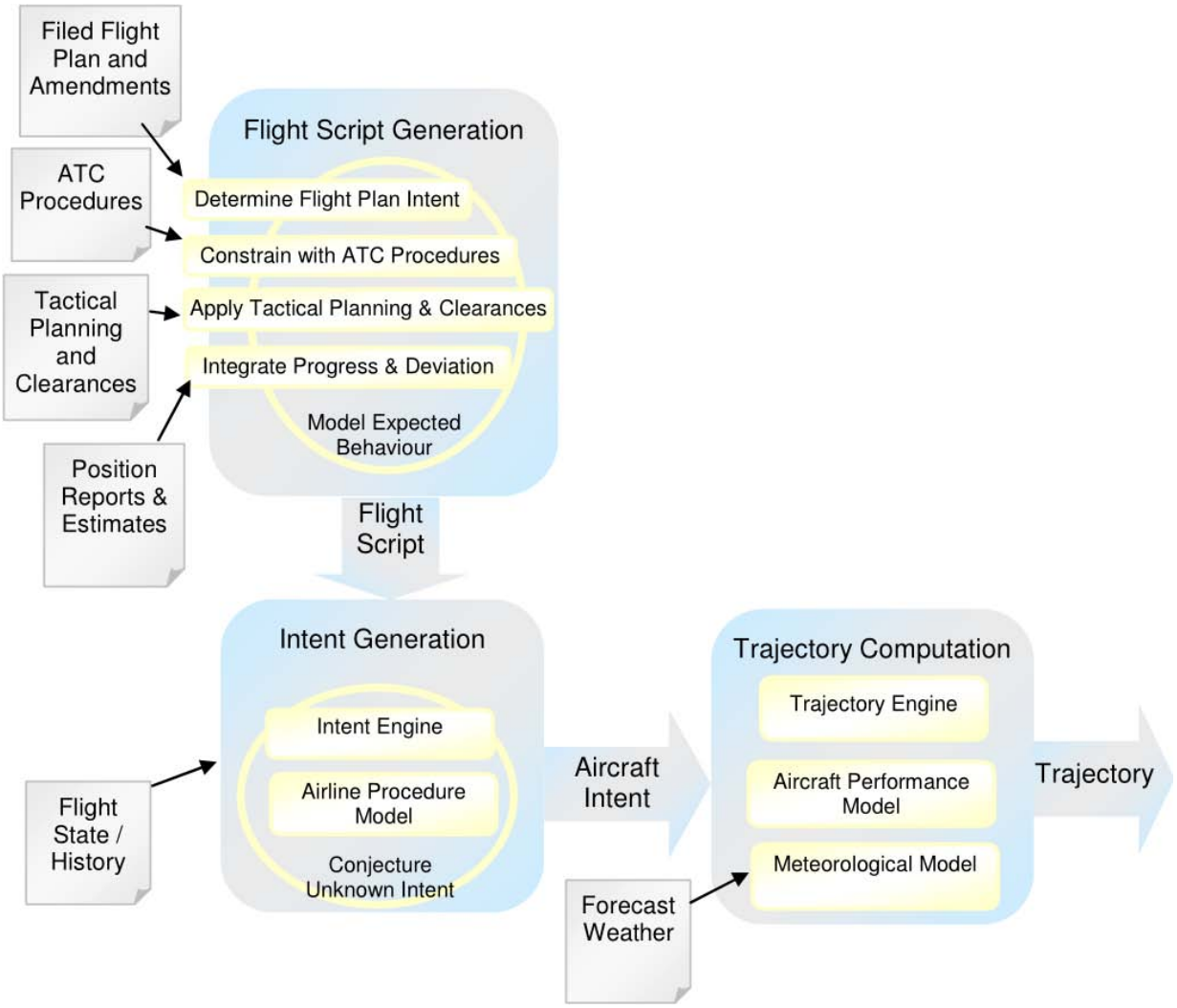
Un prédicteur de trajectoires se décompose classiquement en trois modules organisés en cascade comme illustré par la figure 1.3. Chacun de ces trois modules génère une donnée en sortie, le dernier module renvoie la trajectoire finalement prédite. Le premier module va générer le *flight script* qui décrit les contraintes et préférences associées au vol. Il décrit le vol, les aéroports, les procédures et les contraintes de l'espace aérien. Le second module prend en compte la position courante et les clairances pour spécifier comment le pilote compte mettre en œuvre l'appareil tout en se conformant aux contraintes définies dans le *flight script*. Ce module fournit l'*aircraft intent*. À l'aide de l'*aircraft intent* le troisième module va générer la trajectoire à l'aide d'un modèle météorologique et un modèle des performances de l'avion. Cette thèse se focalise sur ce troisième module, le modèle avion. Le document [FAA / EUROCONTROL 04] propose une classification en 4 classes de ces prédicteurs.

Modèles de classe A

Cette classe modélise la trajectoire ainsi que l'attitude de l'avion. Ceci est fait en modélisant les forces et moments s'appliquant à l'avion. La modélisation des forces est faite en s'appuyant sur les données avions et moteurs fournies par les constructeurs. Ces données sont souvent sous formes de tableaux. Dans ce modèle, l'avion est piloté au travers des commandes des surfaces mobiles et des moteurs, ainsi ce type de modèle nécessite la connaissance des lois de contrôle. Parmi ces prédicteurs, on peut citer BADA 4.0 [Eurocontrol 12a]. Dans BADA 4.0, les forces sont modélisées comme des fonctions paramétriques. Chaque type d'avion a des valeurs de paramètres différentes.

Modèles de classe B

Les modèles de classe B s'appuient également sur une modélisation physique. Toutefois, l'attitude de l'avion n'est pas déterminée et uniquement les forces longitudinale sont utiles i.e. la poussée et la traînée. Une modélisation de ces forces sous forme de tableaux ou de polynôme est alors nécessaire. Ces forces sont alors utilisées pour calculer la puissance disponible qui est ensuite répartie entre l'énergie cinétique et l'énergie potentielle. Parmi ces prédicteurs, on peut citer BADA 3 [Nuic 11]. Dans la famille 3 de BADA, la poussée et la traînée sont modélisées comme des fonctions paramétriques. Chaque type d'avion a des valeurs différentes de paramètres.



Source : Eurocontrol

FIGURE 1.3 – Structure classique d'un prédicteur comme illustré dans [Eurocontrol 10].

Modèles de classe C

Les familles précédentes modélisaient les forces s'appliquant sur l'avion. Dans cette classe, les prédicteurs ne modélisent aucune force. On modélise directement le déplacement. Des fonctions modélisent l'accélération et/ou le taux de montée en fonction de l'altitude par exemple. Parmi ces prédicteurs on peut citer le modèle GAME ([Calders 02]). On peut aussi citer les travaux réalisés dans la thèse [Le Fablec 99] qui utilise des réseaux de neurones artificiels pour modéliser le taux de montée et la vitesse.

Modèles de classe D

Les modèles de classe D sont similaires à ceux de classe C. La seule différence est que le modèle se présente sous forme de tableaux et non de fonctions.

1.4.2 Détection et résolution de conflits

Les outils de détection et la résolution de conflits sont un des éléments pouvant aider le contrôleur dans la gestion d'un trafic croissant. Ces outils doivent prévoir les futures trajectoires car ils visent à anticiper les pertes de séparations et, le cas échéant, à prévoir une série de manœuvres permettant d'éviter ces pertes de séparations. Certains des algorithmes pouvant servir à l'implémentation de ce type d'outils ont été développés par les laboratoires sur le site de l'ÉNAC ([Durand 96, Granger 02]).

Les incertitudes quant à la position future de l'appareil sont modélisées à l'aide de volumes convexes. La trajectoire prédite est modélisée par une séquence de volumes convexes associés à des dates suffisamment proches. Chaque volume est associé à une date et doit contenir la position de l'avion à cette date. Avec cette modélisation illustrée par la figure 1.4, il y a conflit entre deux trajectoires si il existe une date pour laquelle les deux convexes ont une distance latérale inférieure à 5 Nm et une distance verticale inférieure à 1 000 ft.

Si l'on prévoit de donner un ordre à un avion, sa trajectoire future sera différente et donc la séquence de convexes sera aussi différente. La résolution de conflits va consister à choisir un ensemble d'ordres de contrôle à donner aux avions pour que les suites de convexes ne soient pas en conflit tout en limitant le coût des manœuvres effectuées par les avions. Plus il y a d'incertitudes sur la trajectoire future, plus les convexes sont de grande taille et moins il y a de manœuvres résolvant les conflits. Avoir de grandes incertitudes va ainsi réduire l'ensemble des solutions possible S, augmentant mécaniquement [2] le coût de la solution optimale $\min_{s \in S} coût(s)$.

L'espace de recherche étant grand, les algorithmes mis en œuvre dans [Durand 96, Granger 02] ne sélectionnent pas forcément la solution de coût optimale. Prendre en compte un faible nombre d'avions permet de réduire la taille de l'espace de recherche. Pour réduire le nombre d'avions à prendre en compte, on forme une partition en *clusters* des avions étant en conflit avec un autre. Ainsi, préalablement à la résolution de conflits, les convexes

2. On admet ici qu'une incertitude plus grande dans le cas 1 que dans le cas 2 conduit à avoir des ensembles de solutions S_1 et S_2 tels que $S_1 \subset S_2$.

servent à former ces *clusters* d'avions en conflits. Un *cluster* d'avions est défini comme une classe d'équivalence de la relation d'équivalence « est en conflit avec ». Les algorithmes [Durand 96, Granger 02] sont appliqués sur chaque *cluster* séparément. La constitution de ces *clusters* reposent sur les convexes ; ainsi plus les convexes sont grands et plus la taille des clusters augmente, ce qui augmente la taille de l'espace de recherche et rend plus difficile, *a priori*, la recherche de la solution de coût minimum.

En conclusion, les incertitudes sur la trajectoire future vont augmenter le coût des solutions trouvées par les algorithmes de résolution de conflits. Cette augmentation provient d'une augmentation du minimum théorique et d'une augmentation de l'espace de recherche conduisant, vraisemblablement, à la sélection d'une solution plus éloignée de la solution optimale. Le travail effectué dans cette thèse vise à réduire les incertitudes sur les trajectoires futures. On va en particulier s'intéresser à la phase de montée pour laquelle les performances varient grandement entre avions de type différents et même entre avions de même type.

1.5 Cadre de travail

Les travaux récents exploitent l'idée d'un lien permettant l'échange d'informations avancées concernant les intentions de vols. Des travaux se sont ainsi intéressés à la formalisation d'un langage permettant de caractériser une trajectoire de manière non-ambigüe et univoque, mais qui soit suffisamment expressif, pour encoder toutes les trajectoires qu'un avion serait susceptible de choisir ([López-Leonés 07, Konyak 08]). D'autres utilisent les informations calculées par le FMS et qui seraient descendues au sol pour inférer des paramètres physiques comme la masse ([Thipphavong 13a]).

Notre travail se place dans le cadre des systèmes actuellement opérationnels. Dans ce cadre, les informations nécessaires à l'intégration des équations différentielles d'un modèle physique sont manquantes. Pour l'état initial, la masse est inconnue et le système d'équations n'est pas complet ; il faut savoir quelle vitesse, par exemple, va être choisie par le pilote. Dans un tel cadre, des méthodes statistiques passant outre une modélisation physique ont déjà été testées ([Le Fablec 99, Tastambekov 14, Ghasemi Hamed 14]). Ces méthodes n'utilisent nullement un modèle physique, elles prédisent directement une trajectoire. Dans notre travail nous allons utiliser des méthodes statistiques pour pouvoir alimenter un modèle physique. Ce modèle physique mieux renseigné sera comparé à une utilisation plus directe de méthodes statistiques, et à la méthode de référence *i.e.* la prédiction fournie par le modèle BADA utilisé avec les paramètres de référence.

Dans cette thèse, cette comparaison entre modèle physique renseigné et modèle purement statistique porte sur la phase de montée dans l'espace en route. La phase de montée est une phase pour laquelle on observe une grande variabilité dans le taux de montée entre appareils de type différents et même entre appareils de même type.

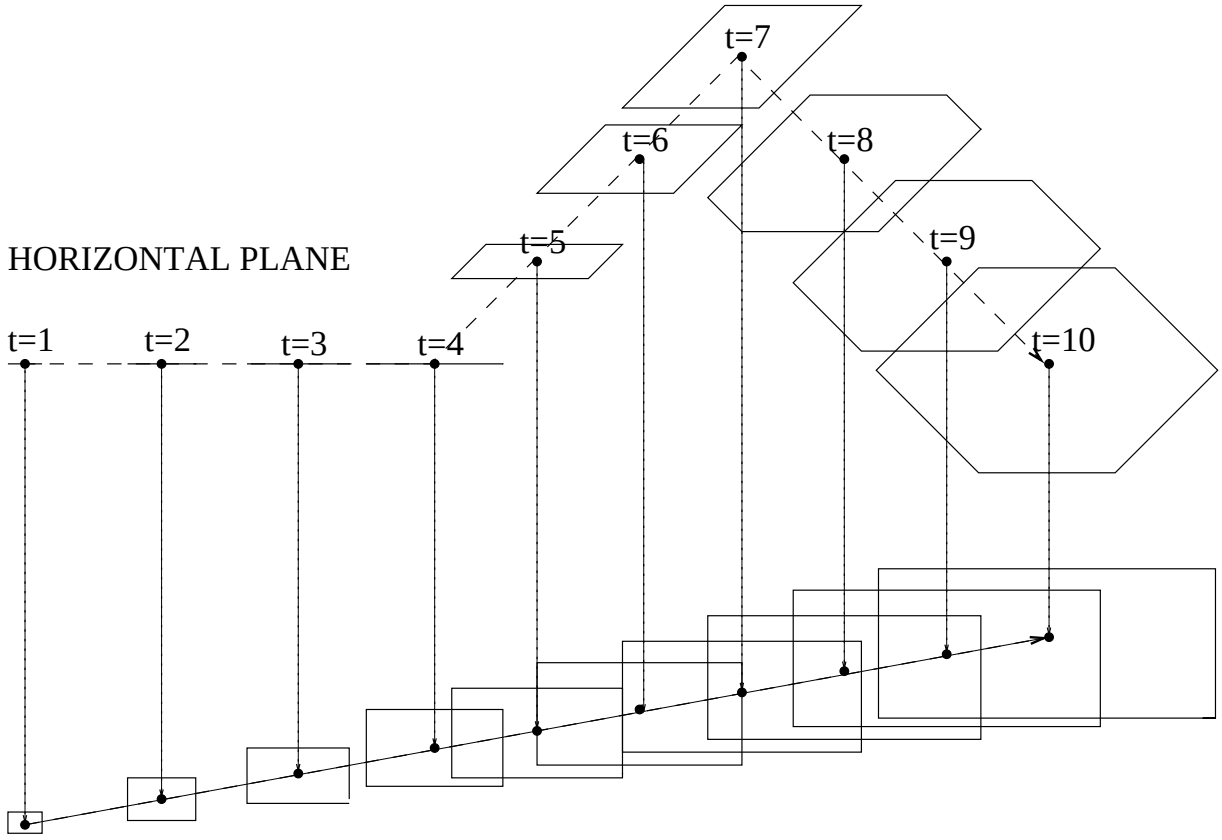


FIGURE 1.4 – Modélisation des incertitudes sur la trajectoire future comme illustré dans [Granger 02].

Chapitre 2

Modélisation physique

Sommaire

2.1 Plusieurs définitions d'altitudes 20
2.2 Modélisation de l'atmosphère 21
2.3 Application de la seconde loi de Newton 22
2.4 Modèle avion BADA 24
2.4.1 Modélisation de la poussée 24
2.4.2 Modélisation de la consommation 25
2.4.3 Modélisation des forces aérodynamiques 25
2.4.4 *Energy Share Factor* 26
2.4.5 Ajustement des coefficients 26
2.4.6 Système d'équations du modèle avion BADA, avec prise en compte du vent 27
2.5 Modélisation BADA des procédures compagnies 28
2.5.1 Profil de poussée pour la montée 28
2.5.2 Profil de vitesse air pour la montée 29
2.6 Calcul de la trajectoire prédite 29
2.6.1 Calcul de la variation d'énergie spécifique 30
2.6.2 Calcul de l'excès de puissance spécifique 31
2.6.3 Calcul du taux de montée et de la vitesse sol 32
2.6.4 Résolution numérique du système d'équations différentielles . . . 33
2.6.5 Importance de l'excès de puissance spécifique 34

Les lois physiques permettent de déduire des relations décrivant le déplacement d'un objet. Par exemple, en considérant un ensemble de forces $F_1, \ldots, F_n$ s'appliquant sur un objet de masse m, la deuxième loi de Newton dit que l'accélération a est égale à la somme des forces sur la masse $\sum_{i=1}^{n} F_i/m$. Il suffit ensuite de modéliser chacune de ces forces F_i en fonction de la position x et vitesse v et de la masse m. Au final, on obtient ainsi l'équation différentielle $a = \sum_{i=1}^{n} F_i(m, x, v)/m$ qui permet le calcul de la trajectoire de l'objet. Avec

l'utilisation des lois physiques, la modélisation de l'accélération se réduit à une modélisation séparée de chaque force qui peut être facilement guidée par une expertise.

Dans ce cadre, ce chapitre présente un modèle à énergie totale couramment utilisé pour la prédiction de trajectoires : BADA. BADA modélise les forces s'appliquant à l'avion et les procédures compagnies guidant la mise en œuvre de l'avion. Ces modélisations peuvent ensuite être exploitées pour construire une trajectoire prédite. Il suffit d'utiliser la deuxième loi de Newton reliant le mouvement de l'avion aux forces s'appliquant à lui.

2.1 Plusieurs définitions d'altitudes

Il est courant d'approximer la surface de la terre par un ellipsoïde. L'altitude géodésique h est la distance entre la surface de l'ellipsoïde et la position de l'avion. On suppose que l'accélération gravitationnelle g n'est que fonction de l'altitude géodésique h, $g = g(h)$. L'altitude géopotentielle H est définie comme solution de l'équation $g(h)dh = g_0 dH$ avec $h = H = 0$ et $g = g_0$ au niveau moyen de la mer.

Toutefois, l'altimètre d'un avion ne calcule pas l'altitude géopotentielle H. Il calcule l'altitude géopotentielle que l'avion aurait s'il volait dans une atmosphère ISA. Cette altitude est l'altitude géopotentielle à laquelle la pression modélisée dans l'atmosphère ISA est égale à la pression mesurée. Pour résumer, cette altitude pression H_p (aussi appelé *altitude pression géopotentielle*) est celle vérifiant l'équation (2.1) avec p la pression et p_{ISA} la pression modélisée par l'atmosphère ISA en fonction de l'altitude géopotentielle H.

$$p = p_{ISA}(H_p) \tag{2.1}$$

Cette altitude pression géopotentielle H_p est calculée à bord puis transmise aux systèmes sol par un transpondeur Mode-C ou Mode-S. C'est cette altitude qui sert à définir les normes de séparation entre les avions. Passée une certaine altitude, l'altitude de transition, l'altitude pression H_p est celle utilisée par les pilotes et le contrôle aérien pour la mise en œuvre des avions. C'est donc cette altitude H_p que l'on essaye de prédire dans cette thèse.

Pour l'utilisation d'un modèle physique, il est nécessaire de relier les différentes variations d'altitude pour pouvoir calculer la variation d'énergie potentielle engendrée par une variation d'altitude pression. Il faut donc relier la variation d'altitude géopotentielle H à la variation d'altitude pression géopotentielle H_p. En calculant la différentielle de l'équation (2.1), on obtient l'équation (2.2).

$$dp = \frac{\partial p_{ISA}}{\partial H}(H_p) dH_p \tag{2.2}$$

Pour aller plus loin il faut écrire dp en fonction de dH. Ceci ne peut se faire qu'au travers d'une modélisation de l'atmosphère.

2.2 Modélisation de l'atmosphère

Pour au moins deux raisons, la prédiction de la trajectoire d'un avion nécessite un modèle atmosphérique. La première est que le système de gestion du trafic aérien et les avions se servent de l'altitude pression [1] comme référence dans le plan vertical ; c'est cette altitude pression que l'on cherche à prédire. La deuxième est liée à la mise en œuvre du prédicteur. Le calcul de la puissance des forces nécessite de connaître certains paramètres de l'atmosphère, comme la densité ρ par exemple. Faute de plus d'informations, l'atmosphère ISA peut être utilisée.

En posant ρ la densité de l'air, R la constante universelle des gaz parfaits et T la température, l'équation des gaz parfaits (2.3) et l'équilibre hydrostatique (2.4) permettent d'écrire l'équation (2.5) reliant la variation d'altitude géopotentielle dH à la variation de pression dp.

$$p = \rho RT \tag{2.3}$$

$$dp = -\rho g \; dh \tag{2.4}$$

$$dp = -\frac{p}{RT} g_0 \; dH \tag{2.5}$$

Ainsi, il suffit de supposer que notre avion évolue dans une atmosphère vérifiant l'équation des gaz parfaits et l'équilibre hydrostatique pour pouvoir écrire l'équation (2.6) reliant variation d'altitude géopotentielle dH à la variation d'altitude pression H_p.

$$-\frac{p}{RT} g_0 \; dH = \frac{\partial p_{ISA}}{\partial H}(H_p) dH_p \tag{2.6}$$

Ces deux équations viennent d'hypothèses standards. C'est d'ailleurs des hypothèses qui sont utilisées dans le modèle ISA de l'atmosphère ([Poles 10a]). Toutefois, ce n'est pas suffisant pour caractériser complètement une atmosphère. Pour ce faire, le modèle ISA rajoute un profil de température défini par l'équation (2.7).

$$dT = \beta_T dH_p \tag{2.7}$$

$$\text{avec } \beta_T = \begin{cases} -6.5 \times 10^{-3} \text{K/m} & \text{si } H_p \leqslant H_{p,trop} \\ 0 & \text{sinon} \end{cases} \tag{2.8}$$

$$\text{avec } H_{p,trop} = 11\,000 \text{ m, altitude de la troposphère} \tag{2.9}$$

Cet ajout permet de caractériser complètement une atmosphère, humidité exclue, en connaissant les conditions aux limites. Pour le modèle ISA, on suppose que les conditions aux limites sont $T(H_p = 0) = 15°C$ et $p(H_p = 0) = 101\,325$Pa. Il est à noter que l'équation (2.7) fait intervenir H_p pour définir le modèle ISA alors que l'altitude H_p est elle-même définie à partir du modèle ISA. Ceci n'est pas gênant car dans le cas du modèle ISA, H et H_p coïncident par définition. Ce jeu d'écritures est utile car le profil de température

1. Par définition de l'altitude pression géopotentielle H_p par rapport à l'altitude géopotentielle H, on a $\frac{dH_p}{T_{ISA}} = \frac{dH}{T}$.

(2.7) sert aussi à définir une famille de modèles d'atmosphère qui diffèrent du modèle ISA uniquement par leurs conditions aux limites. Un élément de cette famille est caractérisé par le couple $(\Delta T, \Delta p)$ défini par les équations (2.10) et (2.11).

$$\Delta T = T(H_p = 0) - T_{ISA}(H_p = 0) \tag{2.10}$$
$$\Delta p = p(H = 0) - p_{ISA}(H = 0) \tag{2.11}$$

D'après l'équation (2.7), cet écart ΔT à l'atmosphère ISA est identique pour toute altitude H_p. Cette famille d'atmosphères paramétrée par $(\Delta T, \Delta p)$ peut servir à simuler des trajectoires dans des atmosphères chaudes ou froides.

Supposer que l'avion vole dans une atmosphère vérifiant l'équation des gaz parfaits et l'équilibre hydrostatique est suffisant pour conclure que $\frac{1}{T}dH = \frac{1}{T_{ISA}(H_p)}dH_p$. Ceci permet surtout d'écrire la variation de l'énergie potentielle $g\frac{dh}{dt}$ en fonction du taux de montée $\frac{dH_p}{dt}$ comme explicité par l'équation (2.12).

$$g\frac{dh}{dt} = g_0\frac{dH}{dt} = g_0\frac{T}{T_{ISA}(H_p)}\frac{dH_p}{dt} \tag{2.12}$$

Pour notre application, nous disposons d'une grille météorologique donnant la température pour différentes altitude H_p. Il n'est pas nécessaire voir souhaitable d'utiliser le profil de température (2.7) qu'utilise la famille de modèles paramétrée par $(\Delta T, \Delta p)$. En utilisant cette grille de température, l'équation (2.12) reste valable en supposant que l'atmosphère dans laquelle évolue l'avion vérifie l'équation des gaz parfaits et l'équilibre hydrostatique.

2.3 Application de la seconde loi de Newton

La plupart des prédicteurs de trajectoires dans les systèmes sol utilisent un modèle à masse ponctuelle pour prédire les trajectoires. Dans un tel modèle, on suppose que toutes les forces s'exercent sur le centre de gravité. Les moments inertiels et accélérations angulaires de l'avion calculés en son centre de gravité sont ignorés. L'avion est modélisé comme une masse ponctuelle de masse m, et l'application de la seconde loi de Newton, donnant l'équation (2.13), nous permet d'écrire l'accélération inertielle $\overrightarrow{a_i} = \frac{d\overrightarrow{V_i}}{dt} = \dot{\overrightarrow{V_i}}$ du centre de gravité comme fonction des forces s'exerçant sur l'avion.

$$m\dot{\overrightarrow{V_i}} = \overrightarrow{Thr} + \overrightarrow{D} + \overrightarrow{L} + m\overrightarrow{g} \tag{2.13}$$

Dans l'équation (2.13), la masse est considérée comme une grandeur stationnaire [2] du point de vue de la dynamique du vol. Toutefois, même si on néglige l'impact de la perte de masse sur le bilan des forces, la consommation du carburant afférente peut et doit être prise en compte lors de l'intégration des équations pour prédire la trajectoire future. Concernant les forces, on suppose que la poussée $\overrightarrow{Thr}$ des moteurs de l'avion s'effectue sur le même axe

2. On suppose que $\frac{d}{dt}(mV_i) = m\dot{V_i}$, ainsi on néglige l'impact de $\dot{m}$ sur l'accélération.

que la vitesse de l'avion dans l'air. La portance $\overrightarrow{L}$, causée par le déplacement de l'aile dans l'air, est perpendiculaire à la vitesse air et dans le plan de symétrie de l'avion. Le vol est supposé symétrique, avec un angle de dérapage nul, il n'y a pas de force aérodynamique orthogonale à $\overrightarrow{V_a}$ et à $\overrightarrow{L}$. Les effets de la rotation de la terre sur la dynamique de l'avion sont négligés (la force de Coriolis est négligée).

Avec les approximations simplificatrices précédentes, et en exprimant les forces dans un système de coordonnées adéquat décrit dans les figures 2.1, 2.2 et 2.3 ; l'équation (2.13) peut se résumer en trois équations scalaires (2.14), (2.15) et (2.16). L'équation (2.14) représente l'accélération en fonction de la poussée, de la traînée, du vent $\overrightarrow{W}$ et de la pente γ_a. L'équation (2.15) représente la variation de cette pente en fonction de la portance et des composantes du vent : composante orientée vers le haut W_Z, composante orientée vers l'Est W_Y et composante orientée vers le Nord W_X. L'équation (2.16) régit la navigation latérale.

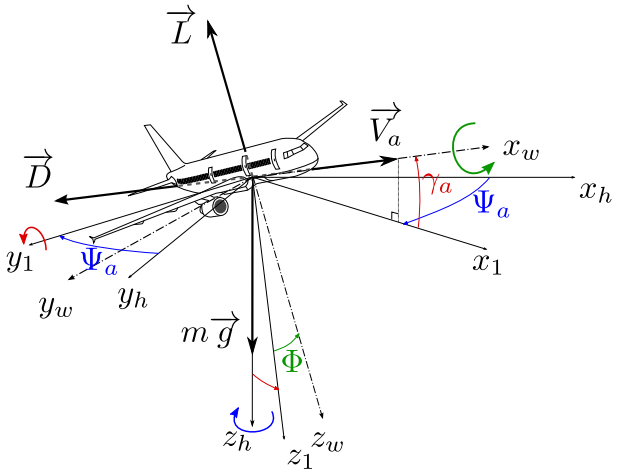


FIGURE 2.1 – Les angles d'Euler attachés à l'avion sont représentés ; Ψ_a est le cap, γ_a est la pente dans la masse d'air et Φ est l'angle de gîte.

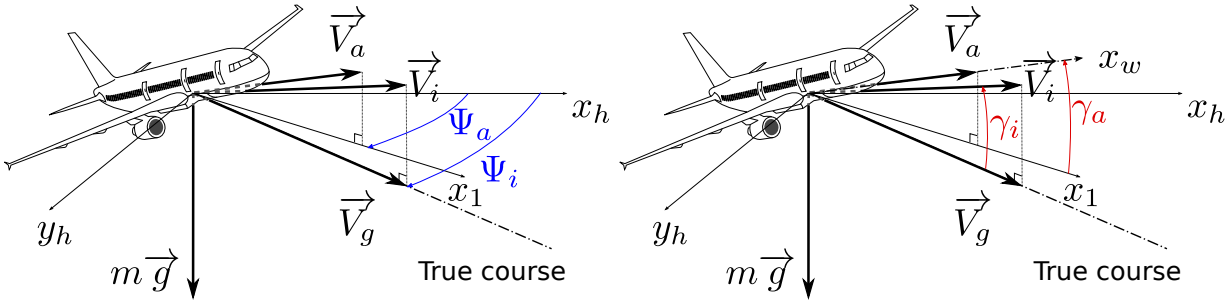


FIGURE 2.2 – Cette figure fait apparaître la vitesse inertielle $\overrightarrow{V_i}$, la vitesse dans l'air $\overrightarrow{V_a}$, la vitesse inertielle dans le plan horizontal $\overrightarrow{V_g}$, le cap Ψ_a, la route Ψ_i, la pente γ_i et la pente dans la masse d'air γ_a (avec x_h pointant le nord).

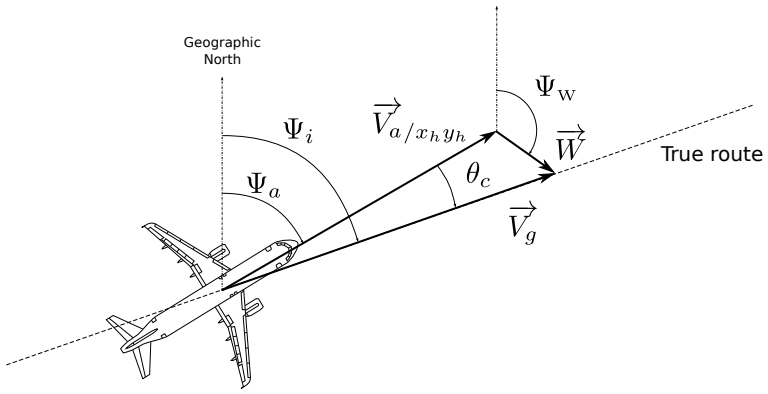


FIGURE 2.3 – Cette figure représente le triangle des vents dans le plan horizontal.

$$\dot{V}_a = \frac{Thr - D}{m} - gsin\gamma_a - \frac{\dot{\overrightarrow{W}}.\overrightarrow{V_a}}{V_a} \tag{2.14}$$

$$\dot{\gamma}_a = \frac{L\cos\Phi}{mV_a} - \frac{g\cos\gamma_a}{V_a} + \frac{1}{V_a}\left[\dot{W}_X \sin\gamma_a \cos\Psi_a + \dot{W}_Y \sin\gamma_a \sin\Psi_a - \dot{W}_Z \cos\gamma_a\right] \tag{2.15}$$

$$\dot{\Psi}_i = \frac{L}{mV_g}(\sin\theta_c \sin\gamma_a \cos\Phi + \cos\theta_c \sin\Phi) - \left(\frac{Thr - D}{mV_g}\right)\sin\theta_c \cos\gamma_a \tag{2.16}$$

2.4 Modèle avion BADA

Le calcul d'une trajectoire nécessite une modélisation des forces s'appliquant à l'avion. La famille 3.X du modèle BADA d'Eurocontrol fournit une modélisation de la traînée D et de la poussée *max climb* $Thr_{\text{max climb}}$. La famille 4.X des modèles BADA, plus précise, existe et est disponible sous licence. Elle couvre un nombre plus restreint de types avion. Les forces sont modélisées plus précisément que BADA 3.9 ([Poles 10b]). Dans ce qui suit, on s'intéresse à la version 3.9 du modèle BADA.

2.4.1 Modélisation de la poussée

Dans le modèle BADA 3.9, la poussée *max climb* est une fonction paramétrique de l'altitude H_p, la vitesse V_a et de l'écart de température au modèle ISA ΔT ([Poles 09]). Cette fonction est définie en deux parties. L'équation (2.17) définit la poussée en atmosphère ISA et l'équation (2.18) définit la poussée en condition non-ISA.

$$(Thr_{\text{max climb}})_{ISA} = t_0 - t_1 H_p + t_2 \frac{1}{V_a} - t_3 \frac{H_p}{V_a} + t_4 {H_p}^2 \tag{2.17}$$

$$Thr_{\text{max climb}} = (Thr_{\text{max climb}})_{ISA} min(max(0.6, t_6 - t_5 \Delta T), 1) \tag{2.18}$$

La poussée *max climb* est la poussée maximum pour la montée. Les poussées d'autres phases de vol sont définies simplement en multipliant la poussée *max climb* par une constante inférieure à 1. À chaque phase est associée sa constante. Cette même fonction paramétrique modélise la poussée de différents types de moteurs : moteur à piston, turbopropulseur ou réacteur. Selon le type de moteurs équipant l'avion, certains coefficients sont supposés nuls. Leur valeur n'est donc pas déterminée par la procédure d'ajustement.

2.4.2 Modélisation de la consommation

Dans le modèle BADA 3.9, la consommation instantanée f_{nom} est considérée comme une fonction paramétrique qui est pour partie proportionnelle à la poussée Thr ([Poles 09]). Cette fonction paramétrique est présentée dans l'équation (2.19).

$$f_{nom} = f_0 - f_1 H_p + \left(f_2 + f_3 V_a - f_4 {V_a}^2\right) Thr \tag{2.19}$$

La consommation nominale f_{nom} est utilisée pour toutes les phases de vol exceptées celles de croisière et de descente au ralenti. Cette même fonction paramétrique modélise la consommation de différents types de moteurs : moteur à piston, turbopropulseur ou réacteur. Selon le type de moteurs équipant l'avion, certains coefficients sont supposés nuls. Leur valeur n'est donc pas déterminée par la procédure d'ajustement.

2.4.3 Modélisation des forces aérodynamiques

Les forces aérodynamiques, L et D, s'expriment en fonction du coefficient de portance C_L et du coefficient de traînée C_D. Ces relations sont explicitées par les équations (2.20) et (2.21), avec S la surface de l'aile.

$$L = \frac{1}{2} C_L \rho S {V_a}^2 \tag{2.20}$$

$$D = \frac{1}{2} C_D \rho S {V_a}^2 \tag{2.21}$$

Le modèle BADA fournit, pour chaque type avion et chaque configuration aérodynamique [3], des valeurs pour la surface S et les coefficients (c_0,c_2). Ces derniers modélisent une polaire Eiffel correspondant à l'équation (2.22).

$$C_D = c_0 + c_2 {C_L}^2 \tag{2.22}$$

Cette modélisation n'est toutefois pas suffisante, on est toujours incapable de calculer les forces aérodynamiques L et D ; il faut une modélisation de C_L. Pour ce faire, on suppose que $\dot{\gamma_a} \backsimeq 0$, ce qui revient à dire que la pente air varie très lentement lors de la mise en

3. Le modèle BADA modélise trois configurations aérodynamiques différentes : une lisse, une d'approche et une d'atterrissage.

œuvre de l'avion. Avec cette hypothèse et en négligeant le vent, l'équation (2.15) donne l'équation (2.23) qui modélise le coefficient de portance C_L.

$$C_L = \frac{2mgcos\gamma_a}{\rho S {V_a}^2 cos\Phi} \underset{cos\gamma_a \backsimeq 1}{\backsimeq} \frac{2mg}{\rho S {V_a}^2 cos\Phi} \tag{2.23}$$

Ce coefficient C_L est normalement une fonction de l'angle d'incidence. C'est cet angle d'incidence que fait varier le pilote pour modifier la portance L et changer ainsi la pente γ_a de l'avion. Ce pilotage de la pente par la variation de la portance est uniquement contenu dans l'équation (2.15). Toutefois, le C_L modélisé par l'équation (2.23) ne peut pas rendre compte de cette dynamique. Ceci est logique puisque ce C_L est obtenu en supposant $\dot{\gamma_a}$ nul dans l'équation (2.15). Il faut rajouter une équation rendant compte du pilotage de la pente.

2.4.4 *Energy Share Factor*

Comme vu dans la sous-section précédente, BADA modélise une portance indépendante de l'angle d'incidence. Ainsi, les équations BADA issues de la seconde loi de Newton sont incapables de rendre compte du pilotage de la pente. On introduit l'équation (2.24) qui est utilisée pour décrire le choix de pente fait par le pilote. Ce choix de pente est spécifié en fixant la valeur de ESF qui correspond à l'*Energy Share Factor*. Cette valeur fixe le partage de la puissance entre la variation de l'énergie potentielle et la variation d'énergie cinétique.

$$g\frac{dh}{dt} = ESF\left(g\frac{dh}{dt} + V_a\frac{dV_a}{dt}\right) \tag{2.24}$$

Spécifier l'ESF n'est qu'une façon parmi d'autre de modéliser le pilotage de la pente. On peut par exemple fixer un profil de vitesse air.

2.4.5 Ajustement des coefficients

Pour un type avion donné, les coefficients du modèle avion BADA sont ajustés en utilisant un jeu de trajectoires. Ce jeu de trajectoires est constitué de 9 trajectoires en montée en condition ISA avec des masses et vitesses différentes. Il est aussi constitué de 3 trajectoires en descente en condition ISA, une trajectoire en croisière en condition ISA et de 4 trajectoires en montée en condition non-ISA. Ces données de trajectoires proviennent du manuel d'utilisation de l'aéronef ou de données de référence générées par des logiciels du constructeur de l'appareil.

Pour chaque trajectoire, on suppose qu'à chaque instant t_i soit connu l'altitude h_i, la masse m_i, la vitesse V_{ai}, la vitesse verticale $\frac{dh}{dt}_i$ et l'ESF ESF_i. Avec ces observations, on est donc capable, à chaque instant t_i, de calculer la vitesse verticale prévue par le modèle avion. De même, on est capable de calculer la consommation prévue par le modèle. Pour une trajectoire, on peut ainsi calculer la somme au carré des écarts. Pour la vitesse verticale, on peut essayer d'ajuster les coefficients minimisant SSE_{ROC}

définie par l'équation (2.25). Pour la consommation, on peut essayer d'ajuster les coefficients minimisant $SSE_{\frac{dm}{dt}}$ définie par l'équation (2.26). L'ajustement des coefficients $(t_0, t_1, t_2, t_3, t_4, t_5, t_6, c_0, c_2, f_0, f_1, f_2, f_3, f_4)$ sur l'ensemble des trajectoires est un problème non-linéaire avec deux critères à minimiser.

$$SSE_{ROC} = \sum_{i=1}^{n} \left[\left(\frac{dh}{dt} \right)_i - \frac{Thr_i - D_i}{gm_i} V_{ai} ESF_i \right]^2 \tag{2.25}$$

$$SSE_{\frac{dm}{dt}} = \sum_{i=1}^{n-1} \left[\frac{m_{i+1} - m_i}{t_{i+1} - t_i} - \left(f_0 - f_1 H_{p_i} + \left(f_2 + f_3 V_{ai} - f_4 {V_{ai}}^2 \right) Thr_i \right) \right]^2 \tag{2.26}$$

Pour traiter ce problème, [Poles 09] propose de considérer des sous-problèmes plus simples. Par exemple, en considérant toutes les trajectoires ISA et (t_5, t_6) fixé, on va ajuster uniquement les coefficients $(t_0, t_1, t_2, t_3, t_4, c_0, c_2)$ pour minimiser les SSE_{ROC}. Puis ensuite, en considérant les trajectoires non-ISA et $(t_0, t_1, t_2, t_3, t_4, t_5, t_6)$ fixé, on ajuste uniquement (t_5, t_6) pour minimiser SSE_{ROC}. Chacun des deux sous-problèmes vus se ramène à une régression linéaire ce qui n'est pas le cas de la recherche de $(t_0, t_1, t_2, t_3, t_4, t_5, t_6, c_0, c_2)$ minimisant SSE_{ROC} sur l'ensemble des trajectoires. Toutefois chacun de ces sous-problèmes fournit un minimum global avec l'hypothèse que certains coefficients sont fixés. Cette hypothèse n'est pas faite dans le problème initial. En itérant la résolution de ces deux sous-problèmes, on améliore $SSE_{ROC}(t_0, t_1, t_2, t_3, t_4, t_5, t_6, c_0, c_2)$ à chaque itération. Pour une description plus complète de cet algorithme, qui en particulier inclus l'ajustement des coefficients $(f_0, f_1, f_2, f_3, f_4)$, on peut se référer à [Poles 09].

2.4.6 Système d'équations du modèle avion BADA, avec prise en compte du vent

On définit l'excès de puissance spécifique P_{es} par la formule $P_{es} = \frac{Thr-D}{m} V_a + gW_Z - \dot{\overrightarrow{W}}.\overrightarrow{V_a}$. Le modèle avion BADA, en tenant compte du vent, est composé de 5 équations. L'équation (2.27) est l'équation du triangle des vents. L'équation (2.28) est équivalente à l'équation (2.14). L'équation (2.29) provient de l'équation (2.16), en supposant $sin\theta_c \backsimeq 0$. L'équation (2.30) modélise la variation de la masse en utilisant l'équation (2.19) modélisant la consommation.

$$\begin{cases} \overrightarrow{V_g} = \overrightarrow{V_a} + \overrightarrow{W} & (2.27) \\ g_0 \dfrac{T}{T_{ISA}} \dfrac{dH_p}{dt} + V_a \dfrac{dV_a}{dt} = P_{es} & (2.28) \\ \dot{\Psi}_i = \dfrac{g_0}{V_g} tan\phi & (2.29) \\ \dfrac{dm}{dt} = -f_{nom} & (2.30) \end{cases}$$

L'équation (2.28) laisse apparaître ici la variation d'énergie spécifique qui vaut par définition $g_0 \frac{T}{T_{ISA}} \frac{dH_p}{dt} + V_a \frac{dV_a}{dt}$. Ce terme est la somme de la variation d'énergie potentielle $g_0 \frac{dH}{dt}$ et la variation d'énergie cinétique $V_a \frac{dV_a}{dt}$. BADA est un modèle à énergie totale. Le système d'équations obtenu est différent de celui de BADA. En effet, dans [Nuic 11] aucun vent n'est pris en compte ; il n'y a aucune référence au triangle des vents et le calcul de l'excès de puissance spécifique est fait sans prendre en compte le vent.

2.5 Modélisation BADA des procédures compagnies

Le prédicteur associé à BADA est un prédicteur de classe B. On modélise uniquement le mouvement du centre de gravité au travers d'une modélisation des forces intervenant dans le calcul de la puissance. Les valeurs de ces forces sont modélisées par des fonctions paramétriques dont on retrouve les paramètres, propres à chaque type d'avion, dans les fichiers BADA.

La mise en œuvre du prédicteur associé nécessite des connaissances, ou à défaut des hypothèses, sur beaucoup de grandeurs. La modélisation des forces intervenant dans le calcul de la puissance est insuffisante pour déterminer un profil de montée. En effet, le calcul de la puissance nécessite de savoir la loi de poussée choisie par le pilote. Une fois ce choix fait, le pilote va faire un second choix en fixant la pente, ce qui revient à fixer la répartition de la puissance entre les variations d'énergie potentielle et d'énergie cinétique. Ces deux choix correspondent à un choix fait sur la mise en œuvre de l'appareil. Ils peuvent être spécifiés par une loi de poussée et une la loi de vitesse. Une fois ceci fait, on peut simuler la trajectoire correspondante à ces lois. Ces lois peuvent être spécifiées de manière indirecte. Par exemple, si on fixe un taux de montée, on impose une variation d'énergie potentielle. Si en plus on fixe la poussée, alors la loi de vitesse est celle qui absorbe la puissance excédentaire P_{es} qui n'a pas été consommée par la variation d'énergie potentielle imposée. On peut se référer à [López-Leonés 07] pour une discussion plus complète sur la modélisation des intentions du pilote concernant le pilotage de son appareil. BADA va justement modéliser des lois de poussée et de vitesse au travers des « procédures compagnies ». On s'intéresse à la phase de montée dans ce qui suit.

2.5.1 Profil de poussée pour la montée

L'excès de puissance spécifique P_{es} est ajusté sur des données issues du constructeur ou du manuel d'utilisation de l'aéronef. BADA définit une seconde modélisation de la puissance excédentaire $P_{es,red}$. Cette modélisation construite avec l'aide de contrôleurs aériens donne des profils plus réalistes ([Nuic 11]) que la puissance excédentaire non réduite P_{es}. Cette puissance réduite introduit un coefficient c_{red} qui est calculé en fonction de l'altitude H_p, la masse courante m, de la masse de base [4] m_{min}, de la masse maximum au décollage [5] m_{max}, de l'altitude maximum h_{max} à la masse m_{max} en condition ISA. Cette réduction dépend

4. Basic Operating Weight.
5. Maximum Take-Off Weight.

des moteurs équipant l'avion. Pour les avions à réaction, on a $k_{red} = 0.15$ et $k_{red} = 0.25$ pour les turbopropulseurs.

$$P_{es,red} = c_{red}\frac{Thr - D}{m}V_a + gW_Z - \dot{\overrightarrow{W}}.\overrightarrow{V_a} \tag{2.31}$$

$$c_{red} = \begin{cases} 1 - k_{red}\frac{m_{max}-m}{m_{max}-m_{min}} & \text{si } H_p < 0.8h_{max} \\ 1 & \text{sinon} \end{cases} \tag{2.32}$$

2.5.2 Profil de vitesse air pour la montée

En aéronautique, il y a plusieurs grandeurs qui peuvent quantifier une vitesse air. La CAS est une vitesse air mesurée par l'avion après correction des erreurs instruments. Elle correspond à la vitesse air V_a nécessaire au niveau de la mer pour créer la même pression dynamique mesurée par l'avion. Le nombre de Mach est le ratio entre la vitesse air V_a et la vitesse du son. Le profil de vitesse BADA est spécifié en fonction de l'altitude H_p. Le profil de vitesse est paramétré par 3 vitesses $(CAS_1, CAS_2, Mach)$ dépendantes du type avion. Pour les avions à réaction, le profil de vitesse est calculé à partir de la vitesse de décrochage V_{stall} pour des altitudes H_p jusqu'à 6 000 ft. Au-dessus de cette altitude, l'avion vole à $min(250 \text{ kts}, CAS_1)$ jusqu'à 10 000 ft. Au-dessus de 10 000 ft, il vole à CAS_2 jusqu'à atteindre l'altitude de conjonction $H_{p,trans}$. Au-delà de cette altitude, l'avion maintient un nombre de Mach égal à $Mach$. Cette altitude $H_{p,trans}(CAS_2, Mach)$ est fonction du couple de paramètres $(CAS_2, Mach)$. C'est l'altitude $H_{p,trans}$ pour laquelle la vitesse air V_a associée à ces deux morceaux se croisent (i.e. $V_a(CAS_2, H_{p,trans}) = V_a(Mach, H_{p,trans})$).

Cette thèse se focalise sur des altitudes supérieures à 10 000 ft. Pour ces altitudes, le choix des paramètres du profil $(CAS_2, Mach)$ résout un compromis fait entre un coût horaire d'exploitation et un coût en carburant. Lorsque l'avion va vite il consomme plus de carburant mais le voyage dure moins longtemps, ce qui permet de réduire le temps d'occupation des ressources que sont l'avion et le personnel navigant. Le choix fait dépend des compagnies.

Faute de connaître le choix fait par l'exploitant de l'avion qui nous intéresse, on peut se reporter aux fichiers BADA qui contiennent des valeurs par défaut, pour chaque type avion, pour les paramètres $(CAS_1, CAS_2, Mach)$ du profil de vitesse. De plus, si l'on ne connaît pas la masse de l'avion, les fichiers BADA contiennent une masse de référence m_{ref} pour chaque type avion.

2.6 Calcul de la trajectoire prédite

On suppose que l'avion évolue dans une atmosphère de température $T = f_T(x, y, H_p, t)$ et de vent $\overrightarrow{W} = \begin{pmatrix} w_X(x, y, H_p, t) \\ w_Y(x, y, H_p, t) \\ w_Z(x, y, H_p, t) \end{pmatrix}$. Ces différentes fonctions peuvent être construites en interpolant la grille météo par exemple. Avec le système d'équations (2.27) à (2.30), une trajectoire est complètement définie en spécifiant l'état initial $(m, V_a, \Psi_i, H_p, x, y, t)$,

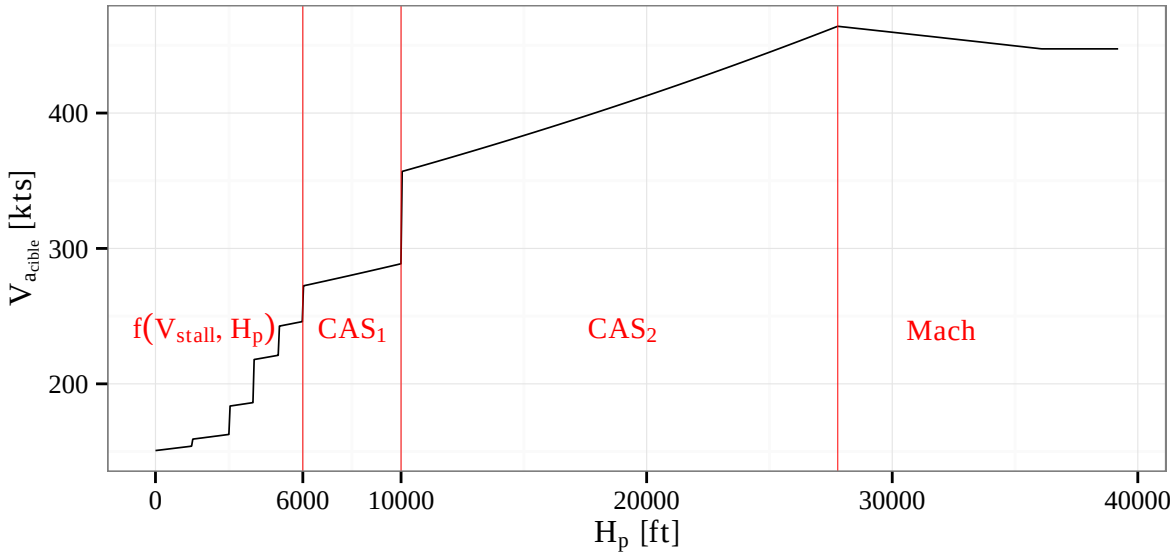


FIGURE 2.4 – Le profil de vitesse air $V_{a_{cible}}$ BADA est spécifié en fonction de l'altitude H_p. Le profil représenté est celui d'un A320 en condition ISA.

le profil de vitesse et le profil de poussée. Le calcul de la trajectoire est fait en utilisant un schéma numérique qui va intégrer les équations du système en tenant compte de l'état initial et des choix de vitesse et de poussée. Pour ce faire il faut calculer les variations des variables de l'état courant. En particulier, la connaissance du taux de montée $\frac{dH_p}{dt}$ et de la vitesse sol dans le plan horizontal V_{gXY} permet de calculer les variations de toutes les variables de l'état courant.

2.6.1 Calcul de la variation d'énergie spécifique

On souhaite exprimer la variation d'énergie spécifique en fonction des variables de l'état courant, du taux de montée $\frac{dH_p}{dt}$ et de la vitesse sol dans le plan horizontal V_{gXY}. On suppose que l'on dispose d'un profil de vitesse air cible $V_{a_{cible}}$ décrit par l'équation $V_{a_{cible}} = f_{V_a}(H_p, T)$. Le profil de vitesse BADA correspond bien à cette hypothèse, la fonction f_{V_a} est paramétrée par $(CAS_1, CAS_2, Mach)$. Suivre ce profil signifie qu'on a $V_a = V_{a_{cible}}$. Ceci permet d'écrire les équations (2.33) et (2.34). Celles-ci expriment $\frac{dV_a}{dt}$ en fonction de l'état courant, de $\frac{dH_p}{dt}$ et de V_{gXY}.

$$\frac{dV_a}{dt} = \frac{dV_{a_{cible}}}{dt} = \frac{\partial f_{V_a}}{\partial H_p}\frac{dH_p}{dt} + \frac{\partial f_{V_a}}{\partial T}\frac{dT}{dt} \tag{2.33}$$

$$\frac{dT}{dt} = \frac{\partial f_T}{\partial t} + \frac{\partial f_T}{\partial H_p}\frac{dH_p}{dt} + \left(\frac{\partial f_T}{\partial x}cos\Psi_i + \frac{\partial f_T}{\partial y}sin\Psi_i\right)V_{gXY} \tag{2.34}$$

Les équations (2.33) et (2.34) servent uniquement à calculer la variation de vitesse permettant de maintenir l'avion sur le profil $V_{a_{cible}}$. Ainsi, il faut que l'avion soit déjà sur le profil $V_{a_{cible}}$ pour que la variation calculée soit pertinente. Toutefois, le profil de vitesse air BADA n'est pas un profil continu par rapport à l'altitude H_p. Or, de par les équations du modèle avion BADA, V_a est nécessairement continue. Ainsi, à la discontinuité de $V_{a_{cible}}$, la vitesse air V_a va différer de $V_{a_{cible}}$. L'avion ne peut donc pas suivre complètement le profil $V_{a_{cible}}$ BADA. Ce problème de discontinuité ne concerne pas l'altitude de conjonction à laquelle il n'y a qu'une rupture de pente, une discontinuité de $\frac{\partial f_{V_a}}{\partial H_p}$.

Ce problème n'est pas explicitement abordé dans le manuel d'utilisation BADA ([Nuic 11]). Il y a toutefois des valeurs d'ESF qui sont suggérées pour pouvoir accélérer ou décélérer fortement au cours de la montée. Dans le cas où l'avion ne peut pas suivre $V_{a_{cible}}$, il est raisonnable de penser à utiliser ces valeurs d'ESF pour modéliser comment l'avion poursuit sa vitesse air cible $V_{a_{cible}}$ lorsque celle-ci n'est pas capturée. Dans le cas où la pente n'est pas spécifiée par une contrainte de vitesse mais par une valeur d'ESF, la variation d'énergie spécifique s'exprime facilement, comme le montre l'équation (2.35).

$$g_0 \frac{T}{T_{ISA}} \frac{dH_p}{dt} + V_{a_{cible}} \frac{dV_{a_{cible}}}{dt} = \frac{1}{ESF} g_0 \frac{T}{T_{ISA}} \frac{dH_p}{dt} \tag{2.35}$$

Au final, que la pente soit fixée au travers d'un profil de vitesse $V_{a_{cible}}$ ou de l'*Energy Share Factor*, la variation d'énergie spécifique est liée linéairement au taux de montée $\frac{dH_p}{dt}$ et à la vitesse sol V_{gXY}. L'équation (2.36) résume ce fait avec a, b et c calculables à partir de l'état courant.

$$g_0 \frac{T}{T_{ISA}} \frac{dH_p}{dt} + V_a \frac{dV_a}{dt} = a \frac{dH_p}{dt} + bV_{gXY} + c \tag{2.36}$$

2.6.2 Calcul de l'excès de puissance spécifique

On doit maintenant calculer l'excès de puissance spécifique. Ceci est relativement aisé si l'on ne prend pas en compte l'effet du gradient de vent $\vec{W}.\vec{V_a}$. Toutefois, le gradient de vent est un paramètre important. Par exemple, on considère un avion à 18 000 ft de type A320 en phase de montée avec une masse m_{ref}, une poussée *max climb* et qui suit le profil de vitesse air BADA. Avec un gradient de vent de 3 kts/1 000 ft, pour cet avion on peut observer une différence de plus de 1 000 ft au bout de 10 min de montée entre la montée

prenant en compte le terme $\dot{\overrightarrow{W}}.\overrightarrow{V_a}$ et celle l'ignorant.

$$\frac{dW_X}{dt}V_{aX} = \left[\left(\frac{\partial w_X}{\partial x}cos\Psi_i + \frac{\partial w_X}{\partial y}sin\Psi_i\right)V_{gXY} + \frac{\partial w_X}{\partial H_p}\frac{dH_p}{dt} + \frac{\partial w_X}{\partial t}\right](V_{gXY}cos\Psi_i - W_X) \tag{2.37}$$

$$\frac{dW_Y}{dt}V_{aY} = \left[\left(\frac{\partial w_Y}{\partial x}cos\Psi_i + \frac{\partial w_Y}{\partial y}sin\Psi_i\right)V_{gXY} + \frac{\partial w_Y}{\partial H_p}\frac{dH_p}{dt} + \frac{\partial w_Y}{\partial t}\right](V_{gXY}sin\Psi_i - W_Y) \tag{2.38}$$

$$\frac{dW_Z}{dt}V_{aZ} = \left[\left(\frac{\partial w_Z}{\partial x}cos\Psi_i + \frac{\partial w_Z}{\partial y}sin\Psi_i\right)V_{gXY} + \frac{\partial w_Z}{\partial H_p}\frac{dH_p}{dt} + \frac{\partial w_Z}{\partial t}\right]\left(\frac{g_0T}{gT_{ISA}}\frac{dH_p}{dt} - W_Z\right) \tag{2.39}$$

Les équations précédentes décomposent [6] le calcul $\dot{\overrightarrow{W}}.\overrightarrow{V_a}$. Elles permettent d'écrire l'excès de puissance spécifique sous la forme d'une équation (2.40) qui est polynomiale du second degré par rapport à $\frac{dH_p}{dt}$ et V_{gXY}, avec a, b, c, d, e et f calculables en connaissant l'état courant.

$$P_{es} = \frac{Thr - D}{m}V_a + gW_Z + a\frac{dH_p}{dt}^2 + b\frac{dH_p}{dt}V_{gXY} + cV_{gXY}^2 + d\frac{dH_p}{dt} + eV_{gXY} + f \tag{2.40}$$

2.6.3 Calcul du taux de montée et de la vitesse sol

Pour calculer le taux de montée $\frac{dHp}{dt}$ et la vitesse sol V_{gXY}, on va utiliser le système d'équations du modèle avion BADA. En particulier, on va utiliser le triangle des vents et l'équation (2.28) reliant la variation d'énergie spécifique à l'excès de puissance spécifique.

$$(2.28) \Leftrightarrow g_0\frac{T}{T_{ISA}}\frac{dH_p}{dt} + V_{a_{cible}}\frac{dV_{a_{cible}}}{dt} = P_{es} \tag{2.41}$$

$$\Leftrightarrow a\frac{dH_p}{dt}^2 + b\frac{dH_p}{dt}V_{gXY} + cV_{gXY}^2 + d\frac{dH_p}{dt} + eV_{gXY} + f = 0 \tag{2.42}$$

Le triangle des vents permet de déduire l'équation suivante :

$$V_a^2 = (V_{gXY}cos\Psi_i - W_X)^2 + (V_{gXY}sin\Psi_i - W_Y)^2 + \left(\frac{g_0T}{gT_{ISA}}\frac{dH_p}{dt} - W_Z\right)^2 \tag{2.43}$$

Ainsi, on obtient deux équations dont chacune est une conique faisant intervenir $\frac{dHp}{dt}$ et V_{gXY}. L'équation (2.43) est une ellipse, tandis que la nature de la courbe associée à l'équation (2.42) est plus dépendante du champ de vent. Résoudre ce système revient à chercher les intersections de ces deux coniques. Ceci peut être fait de manière naïve en « injectant » une équation dans l'autre ce qui conduit à chercher les racines d'un polynôme de degré 4

6. Pour rappel, les composantes W_X et V_{aX} sont orientées vers le Nord, les composantes W_Y et V_{aY} sont orientées vers l'Est et les composantes W_Z et V_{aZ} sont orientées vers le haut.

d'une seule variable. Parmi ces intersections solutions du système d'équations, la contrainte $V_{gXY} \geqslant 0$ permet, on l'espère, d'avoir uniquement une seule solution $\left(\frac{dHp}{dt}, V_{gXY}\right)$.

On peut toutefois éviter de rechercher ces intersections en introduisant deux hypothèses simplificatrices. La première hypothèse est de considérer la composante verticale du vent W_Z comme nulle, ce qui en particulier conduit à avoir $a = 0$ dans l'équation (2.42). En deuxième hypothèse, on suppose que la pente dans l'air est suffisamment faible pour que $cos(\gamma_a) \simeq 1$. Avec ces hypothèses, seule V_{gXY} demeure inconnue dans l'équation (2.43). On peut ainsi aisément déterminer la valeur de V_{gXY}. Une fois ceci fait, il est facile d'en déduire $\frac{dH_p}{dt}$ de l'équation (2.42).

Dans la littérature, ce système de deux coniques n'apparaît pas. En effet, l'effet du gradient de vent n'est pas toujours pris en compte, même si l'article revendique une grande précision comme par exemple [Schuster 12]. Lorsque l'effet du gradient est pris en compte, le terme $\dot{\overrightarrow{W}}.\overrightarrow{V_a}$ est délicat à calculer. En effet, on connaît V_a mais pas Ψ_a et inversement, on connaît Ψ_i mais pas V_g. Ainsi, dans certaines études ([Huchet 06, Mondoloni 06, Gallo 07, Xue 11]), le choix est fait de considérer un vent orienté selon Ψ_i ce qui aboutit à $\Psi_i = \Psi_a$ et rend le calcul de $\dot{\overrightarrow{W}}.\overrightarrow{V_a}$ aisé. D'autres articles ne limitent pas l'étude à des vents orientés selon Ψ_i. Toutefois, ils rajoutent une troisième hypothèse aux deux hypothèses précédentes (i.e. $W_Z = 0$ et $cos(\gamma_a) \simeq 1$). Cette troisième hypothèse, $\Psi_i \simeq \Psi_a$, permet d'aboutir à l'expression utilisée dans [Zhao 96, Slattery 97, Schultz 12]. Le système de deux coniques est équivalent à celui présenté dans [Slattery 97] avant ces trois hypothèses. Contrairement aux deux premières hypothèses, cette troisième hypothèse introduite est contredite par certaines trajectoires de notre jeu de données. On observe parfois des écarts de plus de 15 ° entre Ψ_i et Ψ_a. Un écart de 15 ° peut, suivant l'orientation de $\dot{\overrightarrow{W}}$, engendrer une erreur sur $\dot{\overrightarrow{W}}.\overrightarrow{V_a}$ qui représente jusqu'à 26% de $\|\dot{\overrightarrow{W}}\|V_a$. Sur nos trajectoires, certaines erreurs sur le calcul de $\dot{\overrightarrow{W}}.\overrightarrow{V_a}$ représentent 5% de la variation d'énergie spécifique observée, même pour une variation d'énergie spécifique observée supérieure à 100 W/kg.

2.6.4 Résolution numérique du système d'équations différentielles

Le système d'équations régissant le mouvement de l'avion est un système d'équations différentielles ordinaires. Une résolution analytique de ces équations n'est pas toujours possible. Dans ce cas, on peut employer des méthodes numériques qui vont permettre d'obtenir une approximation de la solution du système d'équations différentielles. On peut trouver une description de ce type de méthodes dans [Hairer 93]. L'approximation de la solution sur $[0; T]$ est décrite par une séquence de valeurs $y_0, \ldots, y_n$ correspondant chacune à la valeur de l'approximation de la solution à des dates différentes. Cette séquence est calculée en utilisant une relation de récurrence obtenue à partir du système d'équations différentielles.

Avec le modèle avion section 2.4.6, le profil de poussée et les procédures compagnies section 2.5.2 vus précédemment, on obtient un système d'équations. Avec ce système, on sait déterminer pour chaque état courant la variation temporelle des variables d'état. Ainsi,

à partir d'un état initial $(m, V_a, \Psi_i, H_p, x, y, t)_{t_0}$, on peut déterminer une suite d'états $(m, V_a, \Psi_i, H_p, x, y, t)_{t_i}$ représentant la trajectoire future. Cette suite d'états constitue une approximation de la fonction solution [7] du système d'équations aux dates t_i. Pour le travail effectué dans cette thèse, on a utilisé une méthode de Runge-Kutta explicite d'ordre 4. Cette méthode a été utilisée pour résoudre la même problématique dans d'autres travaux ([Hadjaz 12, Le Merrer 12]).

2.6.5 Importance de l'excès de puissance spécifique

En notant τ le ratio $\frac{T}{T_{ISA}}$, les équations donnant l'évolution spatiale de l'avion peuvent être mises sous forme matricielle :

$$\begin{pmatrix} \frac{1}{2} & g_0\tau \\ 0 & g_0\tau \end{pmatrix} \begin{pmatrix} \frac{d}{dt}(V_a^2) \\ \frac{dH_p}{dt} \end{pmatrix} = P_{es} \begin{pmatrix} 1 \\ ESF \end{pmatrix} \tag{2.44}$$

En intégrant ces équations de t_0 à t, on obtient la variation d'altitude et de vitesse dans l'intervalle de temps $[t_0; t]$:

$$\begin{pmatrix} {V_a}^2 \\ H_p \end{pmatrix}_t - \begin{pmatrix} {V_a}^2 \\ H_p \end{pmatrix}_{t_0} = \int_{t_0}^{t} \begin{pmatrix} \frac{1}{2} & g_0\tau \\ 0 & g_0\tau \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ ESF \end{pmatrix} P_{es} dt \tag{2.45}$$

L'excès de puissance spécifique provient d'une modélisation des forces s'appliquant à l'avion. Toutefois, l'équation (2.45) permet de voir qu'avoir un bon modèle de chaque force n'est pas nécessaire pour avoir une bonne prédiction de la trajectoire. Il suffit d'avoir un bon modèle de l'excès de puissance spécifique et ce même si la modélisation de chaque force prise séparément n'est pas parfaitement réaliste.

7. On suppose l'existence et l'unicité d'une fonction continue qui serait solution du système d'équations.

Chapitre 3

Apprentissage supervisé et méthodes de régression

Sommaire

3.1 Apprentissage supervisé 36
3.1.1 Notions générales sur l'apprentissage supervisé 36
3.1.2 Évaluation des performances 40
3.1.3 Choix des hyper-paramètres et ensemble de test 41
3.1.4 Réduction du nombre de variables explicatives 42
3.2 Méthodes de régression utilisées 46
3.2.1 Régression linéaire 46
3.2.2 Réseaux de neurones artificiels 52
3.2.3 Gradient Boosting Machine 54

Plusieurs méthodes sont possibles pour calculer la trajectoire prédite d'un avion donnée. Les méthodes utilisées dans nos expérimentations peuvent être classées en deux catégories : les méthodes utilisant l'apprentissage supervisé et celles utilisant un modèle masse-énergie. Le chapitre précédent décrit le modèle masse-énergie. Ce chapitre décrit plus précisément les méthodes d'apprentissage artificiel que nous avons employées. L'apprentissage artificiel recouvre les techniques et méthodes permettant d'extraire de l'information ou un modèle à partir d'exemples issus de l'environnement considéré.

Lorsque les exemples contiennent la réponse que l'on cherche à prédire, on parle d'apprentissage *supervisé*, sinon on parle d'apprentissage *non supervisé*. Dans l'apprentissage *par renforcement*, le modèle appris sert à déterminer l'action à faire. Cette action donne lieu à une récompense déterminée par l'environnement ; à ce titre l'apprentissage *par renforcement* est aussi considéré comme étant de l'apprentissage *supervisé*.

Parmi les problèmes d'apprentissage *supervisé*, la nature de la variable réponse à prédire distingue la *régression* de la *classification*. Prédire une ou des variables réelles correspond à un problème de *régression* tandis que prédire une variable catégorielle correspond à un problème de *classification*.

On distingue également l'apprentissage *en ligne* de l'apprentissage *hors ligne*. Dans le cas de l'apprentissage *en ligne*, ces exemples sont fournis au fur et à mesure. Le modèle appris est mis à jour au fur et à mesure, exemple après exemple. Les méthodes spécifiquement développées pour cette tâche ont pour avantage de pouvoir traiter de grosses bases d'exemples. À chaque nouvel exemple, la mise à jour est faite à un coût constant, indépendant du nombre d'exemples précédemment traités. Le second avantage est de pouvoir suivre une dérive du modèle sous-jacent. En effet, dans ces algorithmes il y a un oubli des anciens exemples qui permet de faire dériver le modèle appris le cas échéant. Dans le cas de l'apprentissage *hors ligne*, les exemples sont considérés tous en même temps pour construire un modèle qui sera considéré comme fixe. Le lecteur intéressé par l'apprentissage artificiel pourra se référer à [Vapnik 99, Hastie 01, Bishop 06, Cornuéjols 10] pour aller plus loin.

Dans notre cas, on a choisi de traiter nos problèmes par le biais de l'apprentissage *hors ligne*. Plus précisément, on va s'intéresser aux problèmes de régression car les variables que l'on cherche à prédire sont des nombres réels.

Une méthode standard pour calculer la trajectoire prédite est d'intégrer les équations d'un modèle masse-énergie. Les modèles masse-énergie requièrent la connaissance d'un grand nombre de paramètres qui ne sont pas disponibles dans les systèmes sol.

Ainsi, à partir d'exemples, on va pouvoir apprendre des modèles qui vont nous permettre de prédire les paramètres manquants d'un modèle physique masse-énergie ou même directement les positions futures.

3.1 Apprentissage supervisé

Dans l'apprentissage supervisé, on utilise un jeu d'exemples $T = (x_i, y_i)_{1 \leqslant i \leqslant n}$ provenant de tirages indépendants issus d'une même loi de probabilité jointe (X, Y), et l'on recherche une fonction h qui permet de prédire y connaissant x. De manière naïve, on souhaite que $y = h(x)$ soit le plus proche possible de nos exemples, et permette de prédire une bonne réponse pour de nouvelles entrées x. C'est un problème classique, largement étudié en statistiques. Dans cette thèse, nous l'abordons par le biais de l'apprentissage artificiel en nous intéressant plus aux aspects pratiques de la prédiction qu'aux propriétés mathématiques des modèles considérés.

3.1.1 Notions générales sur l'apprentissage supervisé

Une fois la problématique de l'apprentissage posée, il nous faut un critère permettant de choisir un prédicteur h parmi un espace d'hypothèses $\mathcal{H}$ contenant tous les prédicteurs candidats. Si l'on sait associer à chaque erreur faite par le prédicteur h une perte, alors on peut choisir h minimisant l'espérance de la perte. En pratique, cette espérance ne peut être calculée directement car la loi de probabilité jointe (X, Y) est inconnue. À défaut de connaître cette loi, on peut utiliser un jeu d'exemples T pour sélectionner l'hypothèse h minimisant la perte sur T : c'est le principe de minimisation du *risque empirique*. Suivant

la définition de la perte, ce principe de minimisation du *risque empirique* peut coïncider avec le principe du maximum de vraisemblance.

On a choisi de s'intéresser au principe de minimisation du *risque empirique* comme principe inductif permettant de sélectionner une hypothèse h. Toutefois, d'autres critères permettant de choisir h existent. Par exemple, Si l'on sait associer à $\mathcal{H}$ une densité de probabilité sur h, alors on peut choisir le prédicteur h le plus probable au regard des données observées. Le principe du maximum a posteriori correspond à cette idée.

Risque réel

Le risque réel $R_{réel}$ est l'espérance de la perte lorsque l'on utilise un prédicteur h. Cette perte est définie au travers de la *fonction de perte* $L \in \mathbb{R}^{\mathcal{Y}^2}$. Elle peut modéliser un coût lié à l'utilisation envisagée du prédicteur.

$$R_{réel}(h) = \mathbb{E}_{X,Y}\left[L\left(h(X), Y\right)\right] \tag{3.1}$$

Le calcul du risque réel requiert la connaissance de la loi de probabilité jointe de (X, Y). Celle-ci étant rarement connue, le risque réel doit être estimé au travers des données observées.

Risque empirique

Le risque empirique $R_{empirique}$ est défini comme la moyenne empirique de la perte calculée sur un jeu de n exemples $T = (x_i, y_i)_{1 \leqslant i \leqslant n}$.

$$R_{empirique}(h, T) = \frac{1}{n}\sum_{i=1}^{n} L\left(h(x_i), y_i\right) \tag{3.2}$$

En supposant que les variables aléatoires $(L(h(X_i), Y_i))_{1 \leqslant i \leqslant n}$ soient indépendantes et suivent une même loi de moyenne et variance finie, on peut appliquer la loi faible des grands nombres. Ceci garantit la convergence en probabilité de $R_{empirique}(h, T)$ vers $R_{réel}(h)$ lorsque la taille du jeu d'exemples augmente. Ainsi, l'utilisation du risque empirique pour estimer le risque réel est justifiée.

Principe de minimisation du risque empirique

Ce qui est intéressant pour l'utilisateur, c'est de choisir h minimisant le risque réel car il correspond à l'espérance de la perte rencontrée lors de l'usage de h. Ce risque réel ne peut pas directement être calculé contrairement au risque empirique. Ces deux grandeurs sont étroitement liées. Pour un jeu d'exemples T donné, l'idée du principe de minimisation du risque empirique est donc de choisir ${h^*}_T$ minimisant $R_{empirique}(h, T)$ en espérant que le risque réel correspondant soit lui aussi minimisé.

Sur-apprentissage

Malheureusement, le risque empirique $R_{empirique}(h^*_T, T)$ et le risque réel $R_{réel}(h^*_T)$ ne convergent pas forcément vers le risque réel minimal sur l'espace d'hypothèses $\mathcal{H}$. En effet, le résultat de convergence vu précédemment s'applique pour une hypothèse h fixée, indépendante de T. Les travaux de Vapnik et Chervonenkis ([Vapnik 91, Vapnik 95, Vapnik 99]) portent, entre autres, sur l'étude de cet écart entre le risque empirique de h^*_T et le risque réel. Ces travaux s'intéressent à des inégalités du type de (3.3) qui permettent de conclure sur la pertinence du principe de minimisation du risque empirique.

$$\mathbb{P}_{X,Y}\text{fixée}, \forall h \in \mathcal{H}, \forall \delta > 0 : \mathbb{P}_T \left[R_{réel}(h) - R_{empirique}(h, T) \leqslant g(\mathcal{H}, \delta, n) \right] > 1 - \delta \qquad (3.3)$$

Dans cette inégalité, le terme $g(\mathcal{H}, \delta, n)$ est un majorant de la différence entre le risque réel et le risque empirique. Ce majorant est croissant avec la capacité d'adaptation de l'espace des hypothèses $\mathcal{H}$. Dans le cas où $\mathcal{H}$ est fini, le terme $g(\mathcal{H}, \delta, n)$ fait intervenir le cardinal de $\mathcal{H}$. De même, dans le cadre de la classification binaire, Vapnik fait apparaître la dimension de Vapnik-Chervonenkis qui est liée à la capacité de $\mathcal{H}$. La capacité de $\mathcal{H}$ désigne sa capacité à exhiber une hypothèse h qui s'ajuste aux observations [1]. Pour un nombre d'exemples n fixé, le comportement des termes $g(\mathcal{H}, \delta, n)$ suggère que plus la capacité de l'espace des hypothèses $\mathcal{H}$ est grande et moins le risque empirique est représentatif du risque réel rencontré.

Cela a des conséquences pratiques sur la mise en œuvre de méthodes d'apprentissage artificiel. Lors du choix de l'espace d'hypothèses $\mathcal{H}$, on doit tenir compte du nombre d'exemples disponibles sous peine de sélectionner une hypothèse h^*_T très performante sur le jeu d'exemples T mais peu performante sur des exemples nouveaux. Ce phénomène, représenté de manière schématique par la figure 3.1, est appelé le *sur-apprentissage*.

Compromis biais-variance

La décision optimale de Bayes h^* définie par l'équation (3.4) est de risque réel R^*. Toute hypothèse h a un risque réel supérieur à R^*. L'hypothèse h^* n'appartient pas forcément à $\mathcal{H}$.

$$h^*(x) = \underset{y \in \mathcal{Y}}{argmin}\ \mathbb{E}_{Y|X=x} \left[L(y, Y) \right] \qquad (3.4)$$

Avec la figure 3.1, on peut voir que le risque réel $R_{rel}(h^*_T)$ n'est pas monotone avec la complexité de l'espace des hypothèses. Pour mieux comprendre ce comportement, on peut décomposer le risque réel comme la somme d'une erreur d'approximation et d'une erreur d'estimation :

$$R_{réel}(h^*_T) - R^* = \underbrace{R_{réel}(h^*_T) - \min_{h \in \mathcal{H}} R_{réel}(h)}_{\text{Erreur d'estimation}} + \underbrace{\min_{h \in \mathcal{H}} R_{réel}(h) - R^*}_{\text{Erreur d'approximation}} \qquad (3.5)$$

1. Exemple : Dans le cas de fonctions polynomiales, la capacité va dépendre du degré du polynôme.

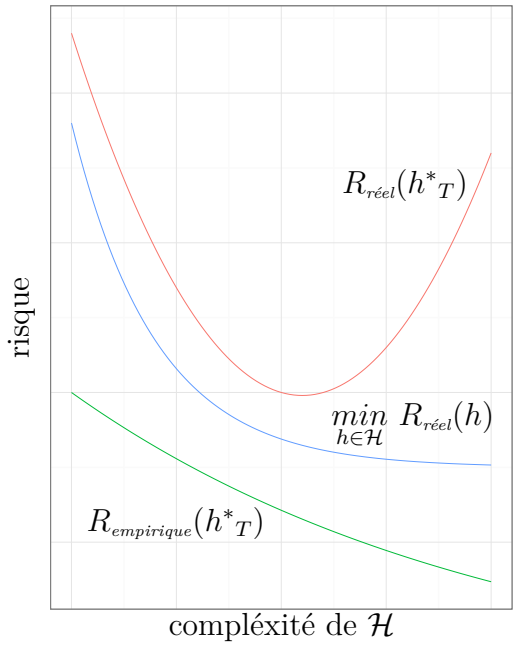


FIGURE 3.1 – On considère un jeu d'exemples T fixé et une succession d'espaces d'hypothèses imbriqués de complexité croissante. Sur ces courbes, l'écart entre $R_{empirique}(h^*_T, T)$ et $R_{réel}(h^*_T)$ augmente avec la complexité de $\mathcal{H}$ jusqu'à atteindre le *sur-apprentissage* : le risque empirique $R_{empirique}(h^*_T, T)$ n'est plus représentatif du risque réel $R_{réel}(h^*_T)$.

L'erreur d'approximation (correspondant au biais) est évidement décroissante[2] avec la complexité de $\mathcal{H}$. La variation de l'erreur d'estimation (correspondant à la variance) avec la complexité est moins claire car l'erreur d'estimation dépend du tirage du jeu d'exemples T. Comme vu pour le *sur-apprentissage*, avec un nombre d'exemples fixe, le risque empirique est moins représentatif du risque réel lorsque la complexité de $\mathcal{H}$ augmente. Par conséquent, la pertinence de ce critère de sélection diminue, entraînant la sélection de mauvaises hypothèses. Ainsi, l'erreur d'estimation augmente avec la complexité de l'espace d'hypothèses $\mathcal{H}$.

Au final, on doit choisir $\mathcal{H}$ optimisant la somme de deux erreurs de monotonie opposées. Dans la pratique, cette somme est décroissante puis croissante avec la complexité de $\mathcal{H}$. Il y a donc un compromis à trouver entre la minimisation de l'erreur d'approximation et la minimisation de l'erreur d'estimation.

Régularisation

La régularisation consiste à modifier le principe de minimisation du risque empirique pour mieux contrôler l'espace d'hypothèses $\mathcal{H}$ sur lequel on recherche h^*_T. Pour ce faire, on va chercher h minimisant le risque empirique sous contrainte que h satisfasse $G(h) \leqslant \mu$ où G est une fonction permettant de pénaliser les hypothèses trop « complexes ». On a

2. On considère des espaces d'hypothèses imbriquées.

ainsi réduit l'espace d'hypothèses effectif. Cette démarche requiert de faire le choix de G et μ. Ce choix traduit un *a priori* sur ce que doit être un bon candidat h ; c'est ce choix qui va conditionner l'espace d'hypothèses effectif.

Dans la mise en œuvre pratique de la régularisation, on se ramène à un problème non contraint dans lequel on a introduit λ, le multiplicateur de Lagrange associé à la contrainte $G(h) \leqslant \mu$. On cherche ainsi h minimisant le *risque empirique régularisé* $R_{reg}(h,T) = R_{empirique}(h,T) + \lambda\, G(h)$.

Par exemple, dans le cas où $\mathcal{H}$ est l'ensemble des modèles linéaires de coefficients w, on peut expliquer la régression Ridge introduite par Hoerl et Kennard ([Hoerl 70]) au travers de cette idée de régularisation. Dans cette méthode, le w estimé est celui minimisant le risque régularisé avec $L(\widehat{y},y) = (\widehat{y}-y)^2$ et $G(w) = \|w\|^2$.

3.1.2 Évaluation des performances

Erreur d'apprentissage

On considère un algorithme d'apprentissage supervisé $\mathcal{A}$. Cet algorithme construit un prédicteur $\mathcal{A}[T]$ à partir d'un jeu de n exemples $T : (x_i, y_i)_{1\leqslant i\leqslant n}$, avec $y \in \mathcal{Y}$ la variable (ou vecteur) à prédire et $x \in \mathcal{X}$ le vecteur de variables explicatives. L'erreur d'apprentissage Err_{app} est le risque empirique sur T lorsque l'on utilise le prédicteur $\mathcal{A}[T]$. Bien souvent, c'est l'erreur que l'algorithme cherche à minimiser lors de la détermination du prédicteur $\mathcal{A}[T]$. Par conséquent, elle a tendance à sous-estimer l'erreur de généralisation qui correspond à la performance effective sur de nouvelles données.

$$Err_{app}(\mathcal{A},T) = R_{empirique}(\mathcal{A}[T],T) \tag{3.6}$$

Erreur de généralisation

L'erreur de généralisation Err_{gen} est le risque réel lorsque l'on utilise le prédicteur $\mathcal{A}[T]$ construit par $\mathcal{A}$. C'est cette erreur que l'on cherche à minimiser au travers de la minimisation de l'erreur d'apprentissage.

$$Err_{gen}(\mathcal{A},T) = R_{réel}(\mathcal{A}[T]) \tag{3.7}$$

Cette erreur ne peut être calculée directement à partir de $\mathcal{A}[T]$; ce calcul nécessite la connaissance de $\mathbb{P}_{X,Y}$. Elle peut néanmoins être estimée en calculant le risque empirique sur un nouveau jeu d'exemples.

Ensemble d'apprentissage, ensemble de validation

Pour pouvoir estimer cette erreur de généralisation, on scinde notre jeu de données en un ensemble d'apprentissage A et un ensemble de validation V. L'ensemble d'apprentissage va être utilisé pour apprendre le prédicteur. L'ensemble de validation va servir à évaluer ses performances sur de nouveaux exemples, une fois le modèle fixé.

$$Err_{val}(\mathcal{A},T,V) = R_{empirique}(\mathcal{A}[T],V) \tag{3.8}$$

En supposant les erreurs indépendantes, $Err_{val}(\mathcal{A}, T, V)$ est un estimateur qui converge en probabilité vers l'erreur de généralisation $Err_{gen}(\mathcal{A}, T)$ lorsque la taille de V augmente.

Pour être utilisé, cette méthode nécessite de disposer de suffisamment de données à partager entre l'ensemble d'apprentissage T et l'ensemble de validation V. Ces deux ensembles doivent être suffisamment grands ; l'ensemble d'apprentissage T doit contenir assez d'exemples pour une bonne estimation du prédicteur par $\mathcal{A}$. De même, plus l'ensemble de validation V est grand et plus l'estimation de l'erreur de généralisation est fiable. Il faut donc T et V grands ; le peu de données disponibles peut conduire ainsi à un compromis à trouver entre la taille de T et de V.

Validation croisée sur k-plis

Dans le cas où peu d'exemples sont disponibles, la validation croisée peut être intéressante. On considère un jeu d'exemples $S : (x_i, y_i)_{1 \leqslant i \leqslant n}$ que l'on partitionne en k parties $(S_i)_{1 \leqslant i \leqslant k}$ de tailles comparables. Par commodité, on pose $S_{-i} = S \setminus S_i$. L'idée de la validation croisée est d'apprendre le prédicteur sur S_{-i} et de tester sa performance sur S_i. Les k résultats obtenus sont ensuite agrégés dans une moyenne pondérée par la taille de S_i. Avec cette procédure, tout les exemples seront utilisés, tour à tour, pour apprendre et évaluer le prédicteur.

$$CV(\mathcal{A}, S) = \sum_{i=1}^{k} \frac{|S_i|}{n} Err_{val}(\mathcal{A}, S_{-i}, S_i) \tag{3.9}$$

Cette procédure peut être relativement coûteuse puisqu'elle requiert l'apprentissage de k prédicteurs. Le prédicteur final dont on cherche à estimer la performance est $\mathcal{A}[S]$; il est obtenu avec plus de données que les $\mathcal{A}[S_{-i}]$. Ceci peut entraîner une estimation pessimiste de la performance, en particulier si la performance de $\mathcal{A}[T]$ progresse rapidement avec la taille de l'ensemble d'apprentissage A. En effet, $CV(\mathcal{A}, S)$ est la moyenne des performances des prédicteurs $\mathcal{A}[S_{-i}]$ obtenus à partir des S_{-i} qui contiennent moins d'exemples que S. Ceci est particulièrement vrai pour k petit.

Dans la validation croisée, on fait la moyenne des erreurs en validation $Err_{val}(\mathcal{A}, S_{-i}, S_i)$. Ainsi, il est difficile de savoir exactement ce qui est estimé au travers de la validation croisée. Si k est grand, on peut penser que les prédicteurs $\mathcal{A}[S_{-i}]$ sont proches de $\mathcal{A}[S]$ et qu'ainsi la validation croisée estime $Err_{gen}(\mathcal{A}, S)$. De même, si k est petit, on peut penser que la validation croisée estime $\mathbb{E}_{S_i}[Err_{gen}(\mathcal{A}, S_i)]$ puisque les prédicteurs $\mathcal{A}[S_i]$ sont entraînés sur des S_i qui peuvent être assez différents. L'erreur la plus intéressante à estimer est l'erreur $Err_{gen}(\mathcal{A}, S)$ puisqu'elle correspond vraiment à la performance avec le prédicteur estimé à partir de toutes les données dont on dispose. Malheureusement, les expérimentations issues de [Hastie 01] sur cette question suggèrent que $CV(\mathcal{A}, S)$ estime $\mathbb{E}_{S_i}[Err_{gen}(\mathcal{A}, S_i)]$ que k soit petit ou grand.

3.1.3 Choix des hyper-paramètres et ensemble de test

Les hyper-paramètres reflètent le choix d'un biais d'apprentissage, d'un espace d'hypothèses $\mathcal{H}$. Ils correspondent donc à un choix *a priori*. Un exemple classique d'hyper-

paramètre est le paramètre de régularisation λ associé à la régression Ridge introduite par Hoerl et Kennard ([Hoerl 70]). Dans la régression Ridge, on estime les coefficients w du modèle linéaire en minimisant l'erreur au carré sous la contrainte $\|w\|^2 \leqslant \tau$. Pour cette contrainte, λ est le multiplicateur de Lagrange associé. On a ainsi restreint l'espace des hypothèses $\mathcal{H}$ en imposant une contrainte sur la norme L^2 des coefficients w du modèle linéaire.

Plus formellement, on considère un algorithme d'apprentissage $\mathcal{A}_\lambda$ paramétré par λ. La sélection de ce paramètre pose problème. Ce choix est difficile à faire *a priori*, sans regarder en détail les données. On ne peut pas le traiter comme un paramètre du modèle linéaire et l'estimer par minimisation de l'erreur d'apprentissage. En effet, ce minimum est atteint en désactivant la contrainte avec $\lambda = 0$, ce qui revient à choisir systématiquement l'espace d'hypothèses le plus grand. On peut contourner ce problème, en choisissant le λ minimisant un estimateur de l'erreur de généralisation, comme l'erreur en validation simple par exemple. Dans l'équation (3.11), le λ choisi est celui minimisant l'erreur en validation sur l'ensemble T_V, appelé *ensemble de test*, en ayant appris à partir de T_T. Ces deux ensembles forment une partition de l'ensemble d'apprentissage T. Avec cette procédure, on définit un nouvel algorithme $\mathcal{A}^*$ (3.10) incluant cette recherche de minimum.

$$\mathcal{A}^*[T] = \mathcal{A}_{\lambda^*}[T] \tag{3.10}$$

$$\text{avec } \lambda^* = \underset{\lambda}{argmin}\ Err_{val}(\mathcal{A}_\lambda, T_T, T_V) \tag{3.11}$$

Les performances de ce nouvel algorithme $\mathcal{A}^*$ sont évaluées de la même manière que n'importe quel autre algorithme.

Pour le choix du λ^*, d'autres critères peuvent être utilisés, comme la minimisation de l'erreur en validation croisée $CV(\mathcal{A}_\lambda, T)$. Dans ce cas, et lorsqu'en plus on évalue la performance de $\mathcal{A}^*$ avec $CV(\mathcal{A}^*, S)$, on parle de *double validation croisée imbriquée*. Il est à noter que le choix du λ peut aussi correspondre au choix de la topologie d'un réseau de neurones, ou du degré du polynôme utilisé dans le cas d'une régression polynomiale.

Dans certains cas, pour le choix des hyper-paramètres λ, il est possible que la comparaison des erreurs de généralisation puisse se faire uniquement à partir de l'erreur d'apprentissage, ce qui permet d'éviter l'emploi d'une procédure coûteuse en temps de calcul telle que la validation croisée. Par exemple, lorsque l'on choisit la log-vraisemblance pour fonction de perte L, et que $\mathcal{A}$ correspond à l'estimateur de maximum de vraisemblance, on peut utiliser le *critère d'information d'Akaike* ([Akaike 74]). Ce critère corrige le biais asymptotique $\lim\limits_{|T|\to\infty} |T|\mathbb{E}_T\left[Err_{gen}(\mathcal{A}, T) - Err_{app}(\mathcal{A}, T)\right]$. D'autres critères utiles à la sélection d'un espace d'hypothèses existent. Parmi les plus courants, on peut citer BIC ([Schwarz 78]) et MDL ([Rissanen 78]) par exemple.

3.1.4 Réduction du nombre de variables explicatives

On considère un jeu d'apprentissage T : $(x_i, y_i)_{1\leqslant i\leqslant n} = (x_T, y_T)$ à partir duquel on souhaite apprendre un prédicteur h qui relie y à x. Comme vu précédemment section 3.1.1,

on peut formaliser ce problème d'apprentissage comme la sélection d'un prédicteur h dans un espace d'hypothèses $\mathcal{H}$. Le principe de minimisation du risque empirique donne un critère de sélection. Toutefois, pour une taille du jeu d'apprentissage T fixée, la pertinence de celui-ci diminue avec la taille de l'espace d'hypothèses $\mathcal{H}$. Or, celle-ci peut dépendre du nombre de variables explicatives. Par exemple, le nombre de poids à ajuster dans un réseau de neurones augmente avec le nombre de variables explicatives. Dans le cas où $\mathcal{X} = \prod_{k=1}^{d} \mathcal{X}_k$, il est possible que certaines composantes de x ne soient pas utile à la prédiction de y et puissent donc être retirées de l'étude.

Typiquement, en retirant une composante à x, on retire de $\mathcal{H}$ tous les prédicteurs qui faisaient intervenir la composante retirée. Ceci est valable par exemple pour la régression linéaire et les réseaux de neurones. Le fait de retirer des composantes inutiles permet de réduire l'espace d'hypothèses $\mathcal{H}$ tout en gardant les prédicteurs les plus performants dans $\mathcal{H}$. Ainsi, en se rappelant du compromis biais-variance, on peut espérer que la performance moyenne des prédicteurs sélectionnés sur l'espace d'hypothèses réduit soit supérieure à celle obtenue sur l'espace d'hypothèses d'origine.

L'idée motivant la sélection de variables est relativement simple ; mais sa mise en œuvre est rendue difficile par la nécessité de faire la distinction entre les variables utiles et inutiles. Cette distinction est délicate à faire car une variable prise individuellement peut sembler peu explicative, mais peut s'avérer très pertinente lorsqu'elle est combinée avec d'autres variables. Cette distinction peut être faite par un expert qui a une idée des variables importantes dans le problème que l'on considère. Hors expertise, on peut utiliser des méthodes qui vont se servir du jeu d'exemples pour déterminer les variables utiles pour résoudre notre problème.

Les méthodes de sélection de variables font partie du processus d'apprentissage, et doivent être inclues à ce titre dans l'algorithme d'apprentissage. Par exemple, si l'on considère T l'ensemble d'apprentissage et V l'ensemble de validation, la sélection de variables doit être basée uniquement sur T et non sur $T \cup V$. Dans le cas contraire, cela peut conduire à une évaluation sur V trop optimiste du prédicteur obtenu ([Hastie 01]). Plus formellement, on considère un algorithme d'apprentissage $\mathcal{A}$ sensible aux variables inutiles que l'on veut rendre plus robuste en rajoutant une méthode de sélection de variables $\mathcal{P}$. $\mathcal{P}[T]$ est la « projection » apprise à partir de T. Cette fonction permet de déterminer un vecteur de variables explicatives de taille réduite à partir du vecteur de variables explicatives original. Au final, l'équation (3.12) décrit l'algorithme $\mathcal{A}^*$ obtenu par la composition de $\mathcal{A}$ et $\mathcal{P}$. Si $\mathcal{P}$ contient des hyper-paramètres, ils deviennent hyper-paramètres de $\mathcal{A}^*$ et peuvent être ajustés par la procédure décrite dans la sous-section 3.1.3.

$$\mathcal{A}^*[T](x) = \mathcal{A}\left[\left(\mathcal{P}\left[T\right](x_T), y_T\right)\right]\left(\mathcal{P}[T](x)\right) \tag{3.12}$$

Ces méthodes sont classées suivant trois approches : les filtres qui sélectionnent les composantes utiles une à une, indépendamment les unes des autres ; les méthodes symbioses sélectionnent un sous-ensemble de variables parmi tout les sous-ensembles possibles ; les méthodes intégrées sont des algorithmes d'apprentissage qui intègrent en eux des sélections de variables, comme les arbres de décision par exemples.

Filtres

Dans la méthode *filtre*, on associe à chaque variable un score supposément représentatif de l'utilité de la variable ; ce score sert à classer les variables. Les variables avec un plus haut score sont supposées être les plus utiles. Comme les variables sont considérées une à une, indépendamment les unes des autres, le calcul des scores est linéaire avec le nombre de variables considérées. Parmi les scores utilisés dans le cas où $\mathcal{X} = \mathbb{R}^d$ et $\mathcal{Y} = \mathbb{R}$, il y a la corrélation linéaire empirique au carré $S(k) = corr^2(x_{\bullet k}, y)$. Ce score $S(k)$ est représentatif de la qualité de l'ajustement obtenue avec un modèle linéaire ayant la composante k comme unique variable explicative. Les méthodes de filtres ne prennent pas en compte l'interaction possible entre les variables. Deux variables fortement corrélées entre elles seront toutes deux choisies si leurs scores sont élevés, ce qui peut poser un problème selon l'algorithme d'apprentissage qui sera utilisé avec les variables sélectionnées. De même, deux variables inutiles séparément peuvent se révéler utiles prises ensembles ([Guyon 03]), mais elles ne seront probablement pas sélectionnées par une méthode *filtre*.

Méthodes symbioses

Dans la méthode *symbiose* popularisée par [Kohavi 97], on considère tous les sous-ensembles que l'on peut former à partir des d variables explicatives ; leur nombre croît en $O(2^d)$. En considérant que l'on utilisera l'algorithme d'apprentissage $\mathcal{A}$ avec le sous-ensemble finalement sélectionné, la méthode *symbiose* cherche le sous-ensemble de variables qui maximise la performance estimée en utilisant $\mathcal{A}$. Cette estimation de performance peut se faire par validation croisée sur T par exemple.

La recherche de ce sous-ensemble est un problème NP-difficile ([Amaldi 98]). Une recherche exhaustive naïve est très coûteuse en temps de calcul ; il faut exécuter au moins 2^d fois l'algorithme d'apprentissage $\mathcal{A}$. Cette recherche peut être éventuellement accélérée par l'emploi d'une stratégie *branch and bound* qui va permettre de ne pas exécuter $\mathcal{A}$ sur tous les 2^d sous-ensembles. Cette idée est exploitée dans [Duarte Silva 01]. Pour rechercher l'optimum global sans toutefois vouloir le prouver, on peut aussi utiliser un algorithme stochastique d'optimisation comme les algorithmes génétiques [Goldberg 89]. Si les méthodes précédentes ne sont pas envisageables, on peut alors utiliser une recherche locale. Parmi celles-ci on peut considérer les méthodes de *hill climbing*. Dans ces méthodes, on a besoin de spécifier la solution initiale et un voisinage de chacun des éléments de notre espace de recherche. Concernant le déroulement de l'algorithme, on considère la solution courante puis on évalue toutes les solutions appartenant au voisinage de la solution courante. Ensuite la meilleure solution du voisinage devient solution courante si elle est meilleure que la solution courante, sinon l'algorithme renvoie la solution courante. Deux implémentations de cette recherche locale sont couramment utilisées : la *forward selection* et la *backward elimination*. Dans la *forward selection*, la solution initiale est l'ensemble vide, et le voisinage de chaque ensemble est constitué de tous les ensembles qui sont obtenus par ajout d'une variable. Dans la *backward elimination*, la solution initiale est l'ensemble des variables lui-même, et le voisinage de chaque ensemble est constitué de tous les ensembles qui sont obtenus par retrait d'une

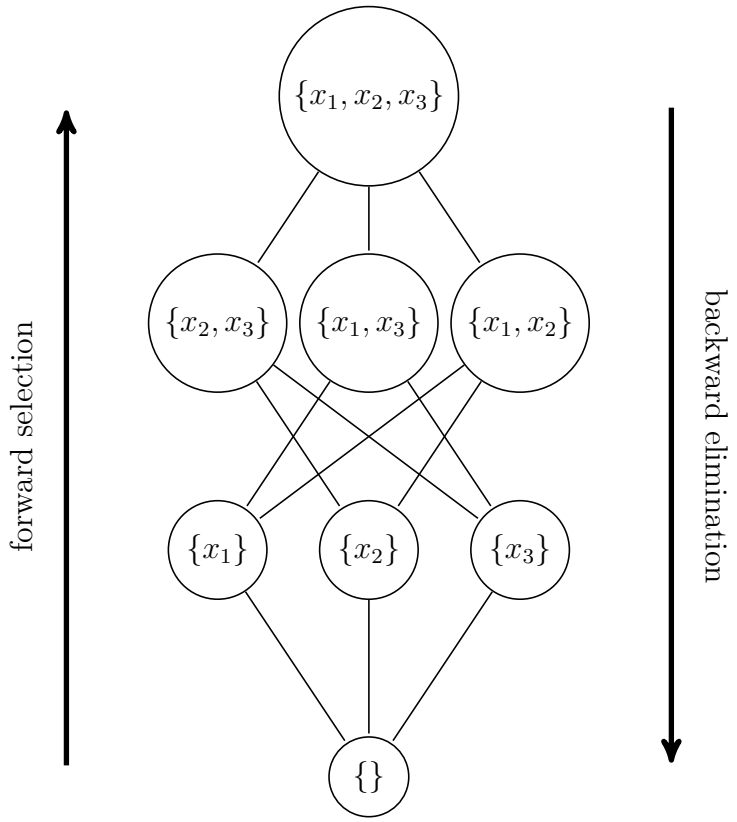


FIGURE 3.2 – Représentation schématique de l'espace de recherche et des voisinages de la *forward selection* et de la *backward elemination* avec $d = 3$.

variable. Une représentation schématique de l'espace de recherche et des voisinages de ces algorithmes est présentée figure 3.2. Comme les méthodes filtres, la *forward selection* a tendance à ne pas sélectionner deux variables inutiles séparément mais utiles ensembles. La *backward elimination* est plus coûteuse que la *forward selection* en calcul car souvent, le temps d'exécution de $\mathcal{A}$ est croissant avec le nombre de variables.

Méthodes intégrées

Ces méthodes sont essentiellement des algorithmes d'apprentissage qui ont un déroulement induisant une sélection de variables. Au final, l'exécution de ces algorithmes fournit un prédicteur et un sous-ensemble de variables qui interviennent dans le prédicteur obtenu. À défaut de fournir un sous-ensemble, elles peuvent aussi associer un score à chaque variable ; contrairement aux méthodes filtres, ce score peut prendre en compte des interactions entre variables. Les méthodes intégrées ont pour avantage de pouvoir prendre en compte des interactions entre variables tout en ayant des temps de calculs plus faibles que ceux des méthodes symbioses.

Parmi ces méthodes, on peut citer par exemple le *gradient boosting machine* ([Friedman 00])

décrit dans la sous-section 3.2.3 et la régression linéaire LASSO ([Tibshirani 94]) qui utilise une régularisation avec la norme L^1 ce qui va contraindre certains coefficients du modèle linéaire à être nuls.

Analyse en composantes principales

L'analyse en composantes principales [3] ([Pearson 01]) permet de réduire le nombre de variables explicatives. Elle diffère cependant des méthodes vues précédemment car elle ne sélectionne pas un sous-ensemble de variables. Elle extrait des variables explicatives en transformant les variables explicatives de départ.

On suppose $\mathcal{X} = \mathbb{R}^d$. On considère x la matrice contenant notre jeu d'exemples. $x_{\bullet k}$ désigne toutes les observations de la variable explicative k. $x_{i\bullet}$ désigne les valeurs des d variables explicatives pour l'observation i. On suppose que les $x_{\bullet k}$ sont de moyennes nulles ; si ce n'est pas le cas on peut aisément les centrer. Dans l'analyse en composantes principales, les variables explicatives transformées sont des combinaisons linéaires des variables de départ. L'analyse en composantes principales est une méthode non-supervisée, elle n'utilise pas la variable à prédire y. Elle construit une matrice orthogonale P de taille d telle que la matrice de covariance $var(xP)$ soit diagonale. Une telle matrice P existe toujours car $var(x)$ est symétrique. Au final, les nouvelles composantes obtenues $(xP)_{\bullet k}$ ne sont plus corrélées. De plus, comme P est orthogonale, on a $\|x_{i\bullet}P\|_2 = \|x_{i\bullet}\|_2$. On peut ordonner les composantes $(xP)_{\bullet k}$ suivant leurs variances. Ainsi lorsque l'on va vouloir ne prendre que r composantes, on prendra celles de plus grande variance pour conserver la variance de nos données. De plus, cet espace vectoriel V de dimension r sur lequel on projette nos données est celui minimisant $\sum_{i=1}^{n} \|x_{i\bullet} - Proj_V(x_{i\bullet})\|_2^2$. La figure 3.3 présente un exemple illustratif de l'analyse en composantes principales.

3.2 Méthodes de régression utilisées

Cette section décrit les méthodes de régression que l'on a utilisées dans cette thèse. Un large spectre de méthodes de régression existe et il est délicat de savoir a priori quelle méthode va être la plus efficace. Les méthodes utilisées dans cette thèse sont celles qui ont fournies de bonnes prédictions à l'issue d'expérimentations préliminaires.

3.2.1 Régression linéaire

La régression linéaire ([Fox 97, Rao 99]) est une méthode couramment utilisée dans des problèmes d'apprentissage. Sa relative simplicité permet de déterminer certaines de ses propriétés statistiques. On présente ici quelques propriétés statistiques de la régression linéaire et sa version régularisée appelée *Ridge* ([Hoerl 70]).

3. Autrement appelée transformation de Karhunen-Loève.

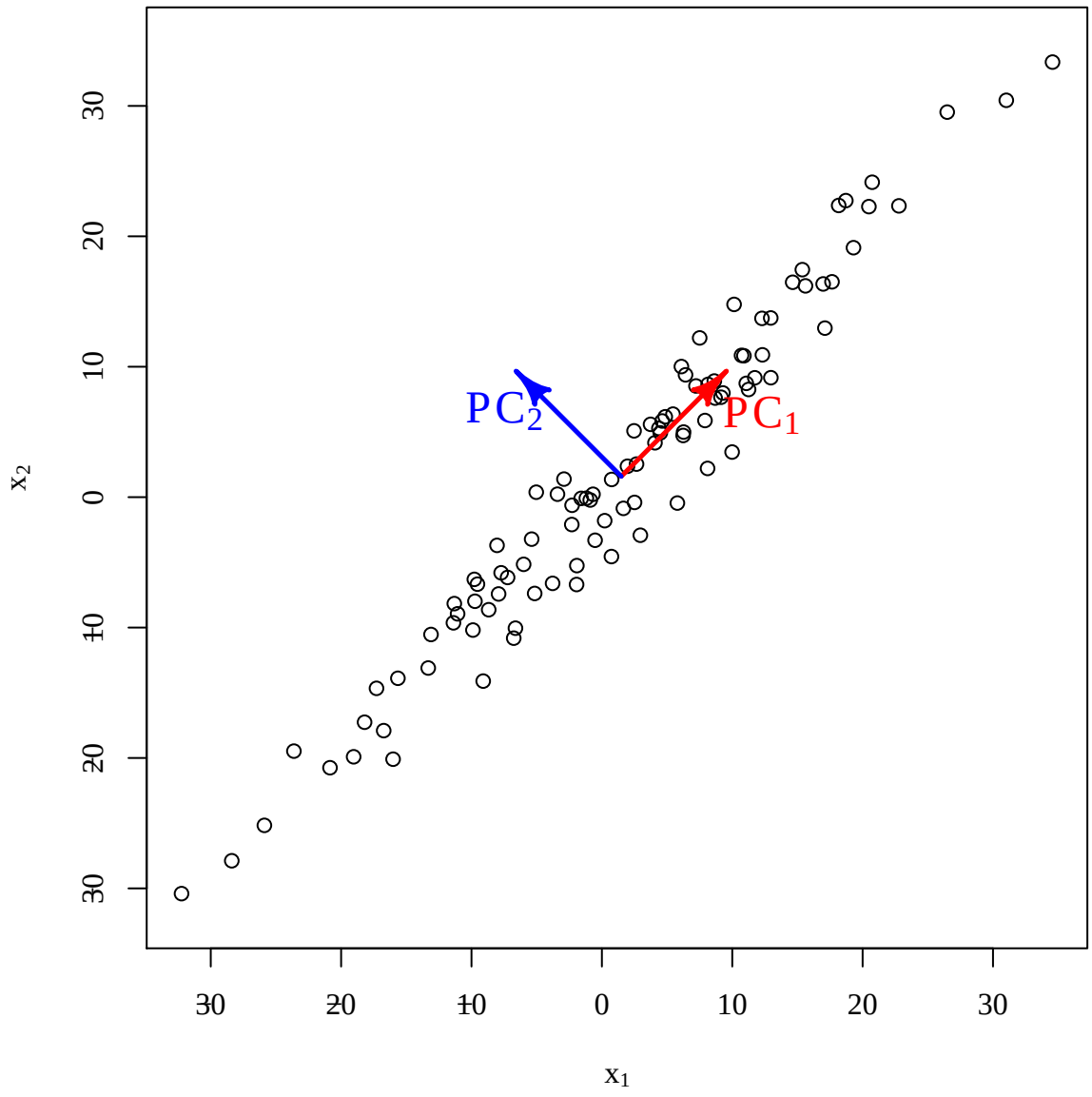


FIGURE 3.3 – Analyse en composantes principales sur un exemple avec $d = 2$; PC_1 et PC_2 forment la nouvelle base orthonormée dans laquelle sont décrites nos observations. PC_1 est associée à la plus grande variance et PC_2 la plus petite.

Estimateur des moindres carrés ordinaires

Avec les notations usuelles, chaque échantillon correspond à $(y_i, x_{i,1}, \cdots, x_{i,p})$, $i = 1, \cdots, n$, avec n le nombre d'échantillons, Y la variable endogène à expliquer, $X = (1, X_1, \cdots, X_p)$ les $p+1$ variables explicatives, et ε les aléas par rapport au modèle. On postule un modèle linéaire paramétré par $a = (a_0, \cdots, a_p)'$ reliant ces variables entre elles. Il reste maintenant à estimer les paramètres de ce modèle à l'aide d'un estimateur statistique. Parmi ces estimateurs, il y a l'estimateur des moindres carrés ordinaires, qui consiste à minimiser la somme des résidus au carré $\sum \varepsilon^2$.

$$y_i = a_0 + \sum_{k=1}^{p} x_{i,k} a_k + \varepsilon_i$$

Formalisons ce problème sous forme matricielle :

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{pmatrix} \begin{pmatrix} a_0 \\ \vdots \\ a_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Avec une notation matricielle plus condensée, on a : $y = xa + \varepsilon$. On note $(.|.)$ le produit scalaire canonique de $\mathbb{R}^{p+1}$.
Notons x' la transposée de x, soit $a \in \mathbb{R}^{p+1}$, on a :

$$\|y - xa\|^2 = \inf_{b \in \mathbb{R}^{p+1}} \|y - xb\|^2$$

$$\Leftrightarrow \|y - xa\|^2 = \inf_{c \in\ Im(x)} \|y - c\|^2$$

$$\underset{\substack{Im(x)\ sous-espace\ vectoriel \\ \|.\|\ norme\ euclidienne}}{\Leftrightarrow} xa = p_{Im(x)}(y), \text{ avec } p_{Im(x)} \text{ projecteur orthogonal sur } Im(x)$$

$$\Leftrightarrow (y - xa) \in\ (Im(x))^{\perp}$$

$$\Leftrightarrow \forall z \in Im(x),\ (z|y - xa) = 0$$

$$\Leftrightarrow \forall z \in \mathbb{R}^{p+1},\ (xz|y - xa) = 0$$

$$\Leftrightarrow \forall z \in \mathbb{R}^{p+1},\ (z|x'(y - xa)) = 0$$

$$\underset{0\ unique\ vecteur\ orthogonal\ à\ tous\ les\ autres}{\Leftrightarrow} x'(y - xa) = 0$$

$$\Leftrightarrow\ x'y = x'xa$$

Il est intéressant de noter que sur des vecteurs centrés, la norme canonique de $\mathbb{R}^n$ est proportionnelle à la variance empirique. Il en est bien évidemment de même pour le produit scalaire canonique et la covariance empirique. Du coup, en supposant y et x centrés, on se rend compte que ce que l'on souhaite minimiser est proportionnel à la variance empirique de ϵ sur l'ensemble d'apprentissage, et que $x\widehat{a}^{mco}$ va correspondre au projeté orthogonal de y sur $Im(x)$ avec la covariance empirique pour produit scalaire.

En supposant H_1 : $x'x$ inversible de rang $p+1$, on a l'estimateur des moindres carrés ordinaires :

$$\begin{aligned}\widehat{a}^{mco} &= (x'x)^{-1}x'y \\ &= (x'x)^{-1}x'(xa+\varepsilon) = a + (x'x)^{-1}x'\varepsilon\end{aligned}$$

On obtient ainsi un estimateur qui va minimiser la somme des résidus au carré. On va maintenant étudier les propriétés de $\widehat{a}^{mco}$ en tant qu'estimateur statistique. Son but étant d'inférer les paramètres a du modèle à partir d'exemples intégrant des aléas modélisés par ε. Pour ce faire, on va devoir considérer différentes hypothèses :

- H_2 : $\mathbb{E}[\varepsilon|X] = 0$
- H_3 : $\mathbb{E}[\varepsilon\varepsilon'|X] = \sigma^2 I_n$, soit des aléas homoscédastiques [4] et non auto-corrélés [5].

Quelques propriétés statistiques sur $\widehat{a}^{mco}$:

1. Estimateur sans biais, sous H_1 et H_2 :

$$\begin{aligned}\mathbb{E}[\widehat{a}^{mco}|X=x] &= \mathbb{E}\left[a + (x'x)^{-1}x'\varepsilon|X=x\right] \\ &= a + \mathbb{E}\left[(x'x)^{-1}x'\varepsilon|X=x\right] \\ &= a + (x'x)^{-1}x'\mathbb{E}[\varepsilon|X] \\ &\underset{H_2}{=} a\end{aligned}$$

2. Variance de l'estimateur, sous H_1, H_2 et H_3 :

$$\begin{aligned}Var[\widehat{a}^{mco}|X=x] &= \mathbb{E}[\widehat{a}^{mco} - \mathbb{E}[\widehat{a}^{mco}])(\widehat{a}^{mco} - \mathbb{E}[\widehat{a}^{mco}])'|X=x] \\ &= \mathbb{E}\left[(x'x)^{-1}x'\varepsilon\varepsilon'x(x'x)^{-1}|X=x\right] \\ &= (x'x)^{-1}x'\mathbb{E}[\varepsilon\varepsilon'|x]\,x(x'x)^{-1} \\ &\underset{H_3}{=} \sigma^2(x'x)^{-1}\end{aligned}$$

3. C'est l'estimateur linéaire sans biais de variance minimum, sous H_1, H_2 et H_3 (Théorème de Gauss-Markov).

Il faut toutefois avoir conscience que x peut correspondre à un système quasi-dégénéré, ce qui impliquera la même chose pour $x'x$. Cela a des conséquences, tant d'un point de vue statistique que numérique.

4. Variance constante
5. Covariance nulle

– D'un point de vue statistique :
La matrice $x'x$ est une matrice positive en tant que produit d'une matrice et de sa transposée, et définie en tant que matrice inversible. On obtient ainsi une matrice diagonalisable $D = Diag(\lambda_1, ..., \lambda_{p+1})$ dans une base orthonormée avec un spectre indicé dans l'ordre croissant inclus dans $\mathbb{R}^{+*}$. On a ainsi, $x'x = P'DP$ avec P matrice orthogonale. Par conséquent, $(x'x)^{-1} = P'Diag(\lambda_1^{-1}, \cdots, \lambda_{p+1}^{-1})P$. On a ainsi :
$Var(\widehat{a}^{mco}|X) = \sigma^2 P'Diag(\lambda_1^{-1}, \cdots, \lambda_{p+1}^{-1})P$

Plus λ_1 sera petit, plus on aura une forte variance selon l'axe $e_1' = P' \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$.

– D'un point de vue numérique :
On doit résoudre $x'xa = x'y$ à l'aide d'un ordinateur ce qui amène forcément des erreurs, la précision avec laquelle on représente un réel étant finie.
Pour étudier dans quelle mesure ces imprécisions dans nos calculs vont impacter la solution finale, on va introduire la notion de conditionnement ([Trefethen 97]). Pour se faire la norme subordonnée à $\|.\|_2$ est intéressante, elle est définie par $\|A\|_{sub} = \sup_{\|v\|_2=1} \|Av\|_2$. On définit maintenant le conditionnement d'une matrice inversible par $\kappa(A) = \|A\|_{sub}\|A^{-1}\|_{sub}$. Dans notre problème, si $x'y$ est non nul, on a les majorations suivantes de l'erreur sur la solution finale :

  – Majoration pour une erreur sur $x'x$:
  $\frac{\|\delta a\|_2}{\|a+\delta a\|_2} \leq \kappa(x'x)\frac{\|\delta(x'x)\|_{sub}}{\|(x'x)\|_{sub}}$, avec $\delta(x'x)$ l'erreur faite sur le calcul de $x'x$ et δa celle faite sur a.
  – Majoration pour une erreur sur $x'y$:
  $\frac{\|\delta a\|_2}{\|a\|_2} \leq \kappa(x'x)\frac{\|\delta(x'y)\|_2}{\|(x'y)\|_2}$, avec $\delta(x'y)$ l'erreur faite sur le calcul de $x'y$ et δa celle faite sur a.

  Or $\kappa(x'x) = \lambda_{p+1}/\lambda_1$, et donc si $x'x$ est quasi-dégénéré, λ_1 sera faible, et $\kappa(x'x)$ sera grand, ce qui rendra la solution finale très sensible aux erreurs de calculs faites sur $x'x$ et $x'y$.

Régression Ridge

Pour faire face aux deux problèmes précédemment évoqués on va utiliser la régression ridge qui va reconditionner le système au prix d'une estimation biaisée. On espère que l'apparition d'un biais soit compensée par une réduction de la variance et de l'impact des imprécisions de calcul. Le problème est formulé de manière différente, on va prendre $x = (x_{\bullet 1}, \cdots, x_{\bullet p})$, on va sortir l'intercept a_0 de a en prenant $a = (a_1, \cdots, a_p)'$. Notre problème devient :
$y = xa + a_0.1_n + \varepsilon$
En posant : $SRC(a_0, a) = \|y - xa - a_0.1_n\|_2^2$, au lieu de prendre (a_0, a) minimisant

$SRC(a_0, a)$, on va minimiser :

$SRC(a_0, a) + \lambda \|a\|_2{}^2$

Avec λ qui va correspondre à une pénalisation de la norme de a. En pratique, elle va minorer les valeurs propres de la matrice que l'on aura inversé, ce qui va garantir un bon conditionnement. En effet, si l'on suppose x centrée/réduite, on a :

$\widehat{a}^{ridge} = {S_\lambda}^{-1} x'y$ et $\widehat{a_0}^{ridge} = \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, avec $S_\lambda = x'x + \lambda I_p$.

On remarque qu'il faudra donc inverser S_λ dont les valeurs propres correspondent à celles de $x'x$ translatées de λ, on améliore donc les problèmes numériques évoqués précédement.

Il reste maintenant à étudier les propriétés statistiques du nouvel estimateur. On a :

$$\widehat{a}^{ridge} = {S_\lambda}^{-1} x'y = {S_\lambda}^{-1} x'x\widehat{a}^{mco}$$

Ainsi, en reprenant les même hypothèses H_1, H_2 et H_3, on aura :

1. Estimateur avec biais, sous H_1 et H_2 :

$$\mathbb{E}(\widehat{a}^{ridge}|X = x) = \mathbb{E}({S_\lambda}^{-1} x'x\widehat{a}^{mco}|X = x) = {S_\lambda}^{-1} x'x\mathbb{E}(\widehat{a}^{mco}|X = x) = {S_\lambda}^{-1} x'xa$$

2. Variance de l'estimateur, sous H_1, H_2 et H_3 :

$$\begin{aligned} Var(\widehat{a}^{ridge}|X = x) &= Var({S_\lambda}^{-1} x'x\widehat{a}^{mco}|X = x) = {S_\lambda}^{-1} x'xVar(\widehat{a}^{mco}|X = x)x'x{S_\lambda}^{-1} \\ &= \sigma^2 {S_\lambda}^{-1}(x'x){S_\lambda}^{-1} \\ &= \sigma^2 P' Diag(\frac{\lambda_1}{(\lambda_1 + \lambda)^2}, \cdots, \frac{\lambda_p}{(\lambda_p + \lambda)^2})P \end{aligned}$$

On a ainsi diminué la variance, et contrairement à l'estimateur des moindres carrés ordinaires, elle est maintenant majorée (par exemple, $\frac{1}{\lambda}$ est un majorant) par un paramètre indépendant de $x'x$.

Régression sur composantes principales

La régression sur composantes principales ([Massy 65]) permet également de remédier aux problèmes causés par la colinéarité des variables explicatives. Elle permet également de contrôler la taille de l'espace d'hypothèses et ainsi d'aider à la résolution du compromis biais-variance. Dans cette méthode on applique une analyse en composantes principales sur les variables explicatives. Comme les composantes principales $(xP)_{\bullet k}$ sont orthogonales entres elles, la méthode *filtre*, évoquée dans la section 3.1.4, utilisant la corrélation empirique $S(k) = corr^2((xP)_{\bullet k}, y)$ rend compte exactement de la réduction d'erreur qu'apporte l'ajout de $(xP)_{\bullet k}$ à un modèle linéaire constitué de composantes principales. Ainsi, les composantes principales sélectionnées seront celles de plus grandes corrélation empirique et non celles de variance maximale. En effet, la variable y peut très bien dépendre de composantes de faible variance comme le souligne [Jolliffe 82]. Le nombre de composantes sélectionnées est un hyper-paramètre permettant de contrôler la taille de l'espace d'hypothèses.

3.2.2 Réseaux de neurones artificiels

Définition

Un réseau de neurones artificiels ([Bishop 95a, Ripley 07]) peut être vu comme une fonction paramétrique. Celle-ci résulte de la composition de fonctions paramétriques élémentaires. Cette succession de compositions est classiquement représentée sous forme d'un graphe orienté comme illustré figure 3.4. Ces graphes orientés acycliques organisés en plusieurs couches ont été popularisés par [Rumelhart 86] qui introduit un algorithme d'apprentissage exploitant cette structuration en réseau. Les sommets contiennent les fonctions élémentaires. Un arc partant d'un sommet correspond au résultat renvoyé par la fonction du sommet considéré. Les arcs arrivant sur un sommet représentent les entrées que la fonction du sommet va utiliser. La couche d'entrée est constituée des variables d'entrées elles-même. Sur la couche de sortie, chaque sommet est associé à une composante du vecteur y que l'on cherche à prédire.

Plus en détails, si un sommet f_θ a pour entrée $(e_1, \ldots, e_r)$, l'arc sortant du sommet renvoie la valeur calculée par la formule 3.13. Les paramètres $(\theta_0, \ldots, \theta_r)$ qui pondèrent les différentes entrées sont les *poids* du réseau de neurones. La fonction f définie 3.14 est identique pour tous les sommets des couches cachées. Cette fonction est appelée *fonction d'activation* ; son choix est généralement fait parmi les fonctions *sigmoïdes*[6] qui sont caractérisées par un comportement linéaire au voisinage de zéro et une saturation lorsque l'on s'éloigne trop de zéro. Pour les problèmes de régression, on prend généralement $g(x) = x$ pour *fonction d'activation* sur la couche de sortie, en lieu et place de f.

$$f_\theta(e_1, \ldots, e_r) = f\left(\theta_0 + \sum_{i=1}^{r} \theta_i e_i\right) \tag{3.13}$$

$$f(x) = tanh(x) \tag{3.14}$$

Ajustement des poids

On considère uniquement un protocole d'apprentissage hors-ligne. Si l'on souhaite ajuster les *poids* du réseau pour minimiser le risque empirique, l'ajustement des *poids* se réduit à un problème d'optimisation. Dans ce cas, l'ajustement des poids peut se faire par exemple, par le biais de méta-heuristiques telles que le recuit simulé ([Kirkpatrick 83]), les algorithmes génétiques ([Goldberg 89]) ou l'évolution différentielle ([Storn 97]). Ces méthodes recherchent un minimum global mais peuvent se révéler coûteuses en temps de calcul. Le réseau de neurones est le résultat de la composition de fonctions élémentaires. Ainsi, il est aisé de calculer le gradient de l'erreur, ce qui rend possible l'emploi de méthodes d'optimisation utilisant cette information. Parmi ces méthodes d'optimisation, on retrouve les algorithmes à directions de descente ([Nocedal 06]) comme la descente de gradient et BFGS ([Broyden 70, Fletcher 70, Goldfarb 70, Shanno 70]) par exemple.

6. Fonctions en forme de « S »

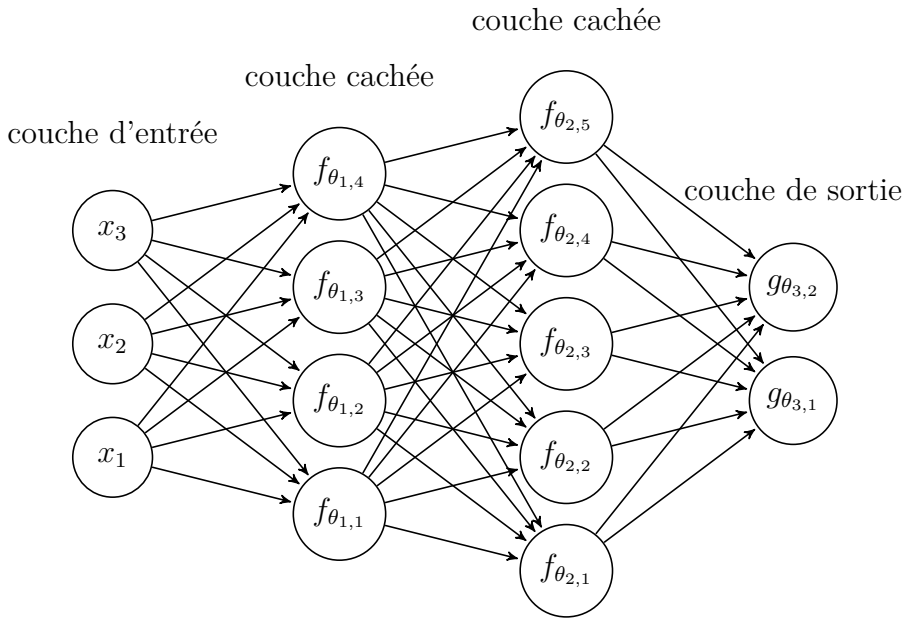


FIGURE 3.4 – Cette figure représente sous forme de graphe un réseau de neurones *feed-forward* à deux couches cachées, 3 entrées et 2 sorties.

Le problème d'apprentissage peut être défini comme la recherche d'un prédicteur minimisant le risque réel. Il peut donc être vu comme un problème d'optimisation. Toutefois, ne connaissant pas la loi jointe (X, Y) sous-jacente, on ne peut pas le minimiser au travers de méthodes d'optimisation classiques car on est incapable de calculer le risque réel. Comme vu dans les sections précédentes, on peut espérer minimiser le risque réel en minimisant le risque empirique. Ce critère peut toutefois conduire à la sélection d'un prédicteur ayant une grande erreur de généralisation. La minimisation du risque empirique n'est pas un critère idéal ; il est d'ailleurs modifié lors de la mise en œuvre de la régularisation.

Ce constat amène à considérer des procédures d'ajustements de *poids* qui ne vont pas obligatoirement sélectionner les *poids* minimisant le risque empirique. Cette idée a été exploitée pour modifier le déroulement d'algorithmes itératifs d'optimisation, [Sietsma 91] propose par exemple de rajouter un bruit aux variables d'entrée à chaque étape de l'algorithme d'optimisation des *poids*. Intuitivement, cet ajout de bruit sur les variables d'entrée va rendre le prédicteur plus « lisse » ; pour tout couple (x_i, y_i) dans T notre ensemble d'apprentissage et tout x proche de x_i, on force $h_T(x)$ à être proche de y_i. Cette technique est étroitement liée aux techniques de régularisation ([Bishop 95b]). Une autre technique, l'*early stopping*, estime à chaque itération de la procédure d'optimisation le risque réel de la solution courante. Le risque réel est supposément décroissant puis croissant avec le nombre d'itérations. Il est donc intéressant de stopper l'ajustement des *poids* quand l'estimation du risque réel croît. Concrètement, l'ensemble d'apprentissage T est coupé en deux ensembles T_T et T_V. L'algorithme d'optimisation utilise uniquement T_T pour ajuster les *poids* et

l'estimation du risque réel de la solution courante est faite avec T_V.

Avant optimisation, les entrées $x_{\bullet k}$ et sorties $y_{\bullet k}$ sont centrées et normalisées. L'ensemble des fonctions paramétrées par les *poids* reste inchangé par ces transformations linéaires des entrées et sorties. Pour les algorithmes d'optimisation ayant besoin d'une solution initiale, cela permet de tirer tout les *poids* dans un même intervalle de taille proportionnelle à $1/\sqrt{r}$ par exemple ([Bishop 95a]). En effet, avec un tel tirage, il est fort probable que l'hyperplan $\theta_0 + \sum_{i=1}^{r} \theta_i x_i = 0$ associé à f_θ traverse le nuage de points formé par les x_i. Ceci permet d'avoir des points des deux côtés de l'hyperplan et suffisamment proches de l'hyperplan pour que $tanh(\theta_0 + \sum_{i=1}^{r} \theta_i x_i)$ ne soit pas saturé pour tous les points x_i. Si tous les points x_i saturent f_θ, le gradient de l'erreur par rapport aux variables θ est proche de 0 ; il y a un « faux plat ». Cette normalisation est donc essentielle au bon déroulement des algorithmes d'optimisation.

3.2.3 Gradient Boosting Machine

Techniques de *boosting*

La théorie de la PAC-apprenabilité [7] développée par [Valiant 84] cherche à caractériser des problèmes de classification binaire. Une classe d'hypothèses est dite PAC-apprenable (au sens fort) si un algorithme d'apprentissage est capable avec une probabilité arbitrairement grande de fournir en un temps raisonnable un prédicteur d'une précision arbitrairement grande. De même, une classe d'hypothèses est dite PAC-apprenable au sens faible si un algorithme d'apprentissage est capable de fournir en un temps raisonnable un prédicteur d'une précision plus grande que celle d'un prédicteur aléatoire, la majeure partie du temps. Étonnamment, ces deux notions sont équivalentes ([Schapire 90]). Pour prouver ceci, [Schapire 90] construit un algorithme démontrant la PAC-apprenabilité au sens fort à partir d'un algorithme la démontrant au sens faible. L'idée pour arriver à ce but est de faire appel à l'algorithme d'apprentissage faible autant de fois que nécessaire sur des distributions à chaque fois modifiées dans le but d'améliorer les résultats. Le prédicteur final est obtenu en agrégeant les résultats des prédicteurs faibles, par un vote majoritaire par exemple.

Le *boosting* repose sur cette idée de construire un prédicteur efficace à partir d'un algorithme faible fournissant un prédicteur fortement biaisé. Typiquement, ce biais est réduit de manière itérative en appliquant l'algorithme faible sur un problème d'apprentissage que l'on modifie à chaque itération en se servant des prédicteurs précédemment appris. Le *boosting* est à distinguer du *bagging* qui combine lui aussi plusieurs prédicteurs dans le but de fournir un meilleur prédicteur. Toutefois, le *bagging* a un impact différent, le *boosting* réduit le biais ; le *bagging* réduit la variance de l'algorithme qu'il utilise en faisant plusieurs apprentissages mais avec des jeux d'exemples différents. À chaque exécution de l'algorithme d'apprentissage, c'est le même problème d'apprentissage qui est considéré pour le *bagging*, contrairement au *boosting*.

7. Probably Approximately Correct

Discrete AdaBoost ([Freund 97]) est l'un des premiers algorithmes exploitant l'idée du *boosting* avec succès pour des problèmes de classification. Plus tard, [Friedman 00] présente une méthode de *boosting* s'inspirant de l'algorithme de descente de gradient qui s'applique aussi bien aux problèmes de régression que de classification. Le *Discrete AdaBoost* peut d'ailleurs être vu comme un cas particulier de cette méthode.

Descente de gradient fonctionnelle

Pour un jeu de n exemples $T : (x_i, y_i)_{1 \leqslant i \leqslant n}$, en considérant une perte L, on cherche à trouver h minimisant le risque empirique sur nos données ou de manière équivalente minimisant la somme des pertes calculées sur chaque exemple.

$$l_n(\widehat{y_1}, \ldots, \widehat{y_n}) = \sum_{i=1}^{n} L\left(y_i, \widehat{y_i}\right) \tag{3.15}$$

$$l(h) = l_n(h(x_1), \ldots, h(x_n)) \tag{3.16}$$

Pour faciliter la compréhension de l'algorithme de descente de gradient fonctionnelle, on décompose cette somme des pertes à l'aide de deux fonctions l_n et l définies par les équations (3.15) et (3.16). On souhaite évidement trouver h minimisant l.

Supposons maintenant que l'on dispose de h prédisant $\left(\widehat{y_i} = h\left(x_i\right)\right)_{1 \leqslant i \leqslant n}$ pour les exemples de T. On est capable de calculer le gradient de l'erreur l_n associé à ces prédictions, comme explicité par l'équation (3.17).

$$g_i = \frac{\partial l_n}{\partial \widehat{y_i}}\left(\widehat{y_1}, \ldots, \widehat{y_n}\right) = \frac{\partial L(y_i, \widehat{y})}{\partial \widehat{y}}\left(\widehat{y_i}\right) \tag{3.17}$$

Ce gradient g nous indique la direction de plus forte pente. Naïvement, avec $\rho > 0$, il faut décaler les prédictions $\widehat{y}$ de $-\rho g$ pour minimiser l_n ; c'est la règle adoptée par l'algorithme de plus forte pente. Les prédictions améliorées $\widehat{y_i}^{+1}$ peuvent alors s'écrire $\widehat{y_i}^{+1} = \widehat{y_i} - \rho g_i = h(x_i) - \rho g_i$. Toutefois, on ne peut pas constituer un nouveau prédicteur avec cette formule car elle ne rentre pas dans un cadre de type « $y = h(x)$ » ; on est incapable d'associer une valeur de g à un nouveau x. [Friedman 00] propose alors d'apprendre un prédicteur g à partir du jeu d'exemples $T_g : (x_i, g_i)_{1 \leqslant i \leqslant n}$. Il reste ensuite à déterminer ρ minimisant $l(h - \rho g)$. Au final, on peut construire un nouveau prédicteur h^{+1} défini par $h^{+1}(x) = h(x) - \rho g(x)$. Comme pour l'algorithme de plus forte pente, une itération ne suffit pas en général ; on peut itérer ce qui précède en considérant h^{+1} en lieu et place de h.

Arbre de régression

Les arbres de régression ([Breiman 84]) sont des modèles qui mettent en jeu des tests successifs pour déterminer une prédiction. Cette succession de tests est représentée sous forme d'arbre binaire. Un arbre binaire représente un partitionnement binaire récursif de l'espace d'entrée $\mathcal{X}$. À chaque nœud, l'espace est partitionné en deux à l'aide d'une condition portant sur les variables d'entrées x.

La valeur prédite par un arbre de régression est celle prédite par sa racine. La valeur prédite par un nœud est celle prédite par son fils de gauche si la condition est vrai ou son fils de droite dans le cas contraire. Ainsi, lorsque l'on souhaite calculer une prédiction, les nœuds de l'arbre ne servent qu'à identifier à quelle partie appartient x. À chaque feuille est associé une partie R_j et un prédicteur γ_j que l'on a appris en utilisant uniquement les exemples appartenant à R_j. Au final, on obtient une partition $(R_j)_{j\in J}$ de $\mathcal{X}$. L'équation (3.18) fournit une écriture du prédicteur h finalement obtenu.

$$h(x) = \sum_{j\in J} \gamma_j(x) I_{R_j}(x), \text{ avec } I_{R_j} \text{ fonction indicatrice de } R_j \tag{3.18}$$

Les prédicteurs γ_j sont généralement choisis constants.

La construction d'un arbre de régression est un problème délicat. Pour une tâche de classification, la construction d'un arbre de décision minimal et cohérent avec un jeu d'exemples donné est un problème NP-difficile ([Hancock 96]). Ainsi, dans la plupart des cas, la construction d'un arbre de régression s'appuie sur une heuristique gloutonne. L'approche la plus populaire est l'approche *top-down* dans laquelle on part de la totalité du jeu d'exemples T. À partir de celui-ci, on détermine le test à appliquer pour que les exemples discriminés par ce test maximisent ou minimisent un critère donné. Une fois le test choisi, on peut appliquer ce qui précède aux deux jeux d'exemples discriminés par le test, T_{vrai} et T_{faux}. L'application de cette procédure à ces deux jeux d'exemples va construire le fils de gauche et le fils de droite.

Pour que le choix d'un test soit fait dans un temps raisonnable, on peut par exemple se restreindre à choisir un test parmi un ensemble prédéfini de tests. Par exemple, on peut considérer un test sur une variable j par rapport à un seuil s conduisant à la création des parties R_1 et R_2 définies par l'équation (3.19). On choisira parmi ces tests possibles celui minimisant un critère, comme celui défini par l'équation (3.20) par exemple.

$$R_1(j,s) = \{(x_i, y_i) \in T | x_{i,j} \leqslant s\} \text{ et } R_2(j,s) = \{(x_i, y_i) \in T | x_{i,j} > s\} \tag{3.19}$$

$$cost(j,s) = \sum_{(x,y)\in R_1(j,s)} (y - \gamma_1(x))^2 + \sum_{(x,y)\in R_2(j,s)} (y - \gamma_2(x))^2 \tag{3.20}$$

D'autres choix sont possibles, par exemple [Heath 93] propose, dans le cadre des arbres de décisions, de séparer les données par le biais d'un hyperplan quelconque. Il en est de même concernant le critère à minimiser ou maximiser ; par exemple on peut choisir un test qui va maximiser l'écart entre les moyennes empiriques de y sur R_1 et R_2 ([Buja 01]).

Cette procédure qui permet à l'arbre de croître est répétée jusqu'à un cas d'arrêt. Ces cas d'arrêts peuvent correspondre à différentes règles. Par exemple, on peut s'arrêter quand l'arbre atteint une certaine profondeur, quand les jeux d'exemples ne contiennent plus suffisamment d'exemples ou quand l'amélioration induite par un nouveau branchement n'est pas significative. Le choix des conditions d'arrêts est important ; si l'arbre est trop petit, il sera peu performant et si l'arbre est trop grand, il aura probablement appris par cœur les exemples.

Pour résoudre ce compromis, l'approche *top-down* est généralement combinée avec une étape dite de *pruning*. Dans cette approche évoquée en premier par [Breiman 84], l'idée

est de réduire l'arbre initialement trop grand en remplaçant des nœuds par des feuilles. Pour ce faire, on calcule pour chaque nœud un score associé à son remplacement. Ce score peut être la perte en précision occasionnée par le remplacement. Dans ce cas, on choisira de remplacer le nœud minimisant cette perte. En itérant cette procédure, on obtient une suite d'arbres partant de l'arbre original jusqu'à un arbre réduit à sa racine. Ensuite, on peut choisir lequel garder parmi cette suite d'arbres en estimant les performances de ces arbres sur un jeu d'exemples non utilisé. C'est cette approche qui est mise en œuvre dans le *weakest link pruning* ([Breiman 84]).

L'implémentation la plus employée des arbres de régression considèrent des tests de type $x_j \leqslant s$ et des prédicteurs γ_j constants. Ces arbres de régression ont plusieurs avantages. Les arbres de régression sont insensibles aux transformations strictement croissantes des variables d'entrées. Utiliser x_j, $log(x_j)$ ou $exp(x_j)$ conduit au même résultat. Un corollaire de cette remarque est la robustesse des arbres de régression aux *outliers* sur les variables d'entrée. De plus, en utilisant un arbre de petite taille, la construction d'un arbre de régression conduit naturellement à une sélection de variables. Ces arbres de régression peuvent gérer de manière élégante les valeurs manquantes ([Breiman 84]). Ils peuvent aussi inclure des variables catégorielles. Ils ont toutefois des inconvénients. Ils sont généralement peu performants pour la prédiction. La construction d'un arbre de grande taille est généralement instable ; une faible variation du jeu d'exemples peut entraîner une grande variation dans l'arbre obtenu. Alors que l'extraction d'une combinaison linéaire est aisée pour la régression linéaire et les réseaux de neurones, elle est toutefois délicate pour les arbres de régression car les tests ne font intervenir qu'une variable.

Gradient Boosted Trees

Dans [Friedman 00], la descente de gradient fonctionnelle est employée avec des arbres de régression. Il est à noter que la descente de gradient fonctionnelle peut s'employer avec d'autres méthodes de régression comme la régression linéaire par exemple ([Bühlmann 07]). Dans nos expérimentations, nous avons utilisé l'implémentation disponible dans le package R gbm ([Ridgeway 07]). Cette implémentation est identique à la description faite dans [Hastie 01]. Elle emploie des arbres de régression avec des tests de type $x_j \leqslant s$ et des prédicteurs γ_j constants. Ces arbres sont construits en utilisant le *weakest link pruning* ([Breiman 84]).

Des arbres avec un faible nombre de feuilles J sont utilisés pour apprendre les gradients successifs. Avec m itérations de l'algorithme de descente de gradient fonctionnelle, on obtient le prédicteur h décrit par l'équation (3.21) comme la somme des arbres de régression h_i. Chacun de ces arbres fait intervenir au plus $J-1$ variables. Ce nombre de feuilles J permet de contrôler le niveau d'interaction entre les variables. Par exemple, si l'on sait que chaque variable a une contribution dans y indépendante des autres variables, on va pouvoir fixer $J=2$. Cette possibilité de contrôler le niveau d'interaction peut améliorer la performance du prédicteur [Hastie 01].

$$h(x) = \sum_{i=0}^{m} h_i(x_{j_{1,i}}, \ldots, x_{j_{J-1,i}}) \tag{3.21}$$

Pour éviter de sur-apprendre, on peut envisager une stratégie de *early stopping* qui correspond ici à stopper la descente de gradient fonctionnelle quand une estimation du risque réel croît. Cette sélection peut aussi être faite *a posteriori*. On choisit le prédicteur correspondant à l'itération m^* minimisant l'estimation du risque réel. D'autres stratégies sont possibles, on peut en plus envisager d'introduire un *learning rate* ν dans la formule $h^{+1}(x) = h(x) - \rho g(x)$. Cette formule devient $h^{+1}(x) = h(x) - \nu\rho g(x)$, avec $0 < \nu < 1$, et permet d'améliorer la performance du prédicteur obtenu ([Friedman 00]). En plus de ces deux hyper-paramètres, [Friedman 02] introduit un autre raffinement améliorant les résultats et le temps calcul lors de l'apprentissage. À chaque itération, un sous-ensemble du jeu d'exemples est tiré, sans remise. Le gradient g est appris en utilisant ce sous-ensemble d'exemples. Il est a noté que cette modification n'est efficace que conjuguée avec un *learning rate* faible.

Chapitre 4

Données de trajectoires

Sommaire

4.1 Obtention des exemples 59
4.1.1 Types de données 59
4.1.2 Deux jeux de données 61
4.1.3 Lissage des données 62
4.1.4 Filtrage des trajectoires 72
4.1.5 Construction des exemples 72
4.2 Analyse des exemples 72
4.2.1 Profils de température observés 74
4.2.2 Profils de vitesse observés 74
4.2.3 Étude de la variation d'énergie spécifique 77
4.2.4 Profils d'Energy Share Factor 84
4.3 Conclusion 87

Ce chapitre décrit l'obtention, à partir des trajectoires d'avions, des jeux d'exemples que l'on utilise. Tout notre travail est basé sur ces données ; ainsi cette étape est essentielle au bon déroulement de cette étude. Un soin particulier a été apporté au lissage des trajectoires et ceci dans l'idée d'estimer des dérivées temporelles. Celles-ci vont permettre le calcul de grandeurs physiques qui sont essentielles. Ces dernières sont aussi modélisées par BADA. Ainsi, on a pu mener une analyse comparative entre les valeurs effectivement prises par ces grandeurs et les valeurs issues de la modélisation qui en est faite dans BADA.

4.1 Obtention des exemples

4.1.1 Types de données

Les données utiles pour la prédiction de trajectoires sont de différentes natures et viennent de différentes sources. Dans cette sous-section, on décrit les données servant à la construction de nos jeux de trajectoires.

Données radar

Les radars récupèrent la position des avions dans le ciel à intervalles réguliers. Il y a deux types de radars, primaire ou secondaire :

- Le radar primaire émet une onde électromagnétique. Cette onde va être réfléchie par l'avion, ce qui va permettre au radar de le situer en azimut et en distance. Ce type de radar n'apporte aucune information sur l'altitude. Il existe des radars primaires donnant l'angle d'élévation mais ils sont plus utilisés dans le domaine militaire que civil.
- Le radar secondaire interroge un équipement de l'avion, le transpondeur. Cet équipement émet en réponse un signal, ce qui va permettre au radar de le situer en azimut et en distance. L'information contenue dans ce signal dépend des capacités du transpondeur et du mode sélectionné sur celui-ci.
  Si le Mode-C est sélectionné sur le transpondeur, le signal contient l'altitude pression mesurée à bord et le code transpondeur qui permet d'identifier l'avion. Cette altitude pression est d'abord numérisée avec une granularité de 100 ft puis transmise.
  Le Mode-S est une amélioration du Mode-C. Comme pour le Mode-C, le signal peut contenir des grandeurs issues des équipements à bord de l'avion. Dans le cas du Mode-S ELS , les grandeurs transmises sont identiques au Mode-C. Dans le cas du Mode-S EHS, le transpondeur transmet d'autres grandeurs en supplément comme la vitesse air et l'angle d'inclinaison par exemple. Dans tous les cas, l'altitude pression mesurée est numérisée plus finement avec une granularité de 25 ft.

Plusieurs radars peuvent mesurer la position d'un même avion. Les différentes mesures concernant un même avion sont fusionnées pour donner une unique succession cohérente de plots radar.

Données plan de vol

Les données COURAGE, dont on peut voir un extrait figure 4.1, contiennent les plans de vol des avions. Un plan de vol décrit la route que l'avion a prévu de suivre. Il contient notamment l'aéroport de départ et celui d'arrivée, le niveau de vol et la vitesse de croisière souhaités. Pour un même vol, il existe trois versions d'un plan de vol COURAGE, chaque version correspond à un moment différent de la vie du vol. Il y a la demande initiale faite une semaine avant le vol, la demande finale faite une heure avant le vol et le plan de vol réalisé qui correspond au vol effectivement réalisé.

Données météorologiques

Les données météorologiques sont issues du modèle de prévision ALADIN opérationnel à Météo-France. Ce modèle est couplé au modèle ARPEGE qui couvre l'ensemble du globe avec une résolution variable. Le modèle ALADIN est constitué d'une grille couvrant toute la France. Les mailles de cette grille sont espacées de 0.1 degré en latitude pour les mailles Est/Ouest et en longitude pour les mailles Nord/Sud. Verticalement, les grilles sont situées sur des isobares qui ne sont pas espacées de manière régulière. Toutes les 6 heures, les

```
20 AFR2142 LFPG LSGG 8748 0 A319 0
21 720 270 450
22 0 N N 0
31 LPG WDG PG264 PG289 PEKDG PEKIM TFL4 PILUL TD11X PIXIS ATN LUSAR SAUNI
32 735 736 736 739 746 746 747 748 748 756 759 764 765
33 0 22 34 85 200 270 270 270 270 270 260 230 230
34 7
41 RD DG UP UT IG
42 718 722 722 748 750
43 735 742 743 758 765
44 766
45 1 3 6 11 13
46 4 7 10 12 13
```

FIGURE 4.1 – Extrait d'un fichier de plans de vols COURAGE. La ligne 20 donne des informations sur le vol : identifiant (AFR2142), origine (Paris-CDG LFPG), destination (Genève LSGG), modèle d'avion (A319) ; la ligne 21 décrit l'heure de départ, le niveau de vol et la vitesse de croisière souhaités.

prévisions météorologiques sont mises à jour. Concernant nos fichiers, on a 10 isobares ; l'isobare la plus basse est l'isobare 1 000 hPa et la plus haute est l'isobare 250 hPa. On dispose de deux échéances de prévisions, l'échéance zéro et l'échéance à 3 heures. L'échéance zéro correspond aux grandeurs prévues en chaque nœud de la grille à la date de prévision. L'échéance à 3 heures correspond aux grandeurs prévues en chaque nœud de la grille à la date de prévision plus 3 heures.

4.1.2 Deux jeux de données

Trajectoires Mode-C

Les données de trajectoires Mode-C sont issues d'enregistrements du Système de Traitement Radar (STR) du Centre en Route de la Navigation Aérienne (CRNA) Nord. On dispose de deux mois d'enregistrements, juillet 2006 et janvier 2007. La figure 4.2 est un extrait de ces enregistrements. Les grandeurs associées à chaque plot ont subi une numérisation, comme illustré figure 4.3. De même, la datation des plots est faite à la seconde près. Cette granularité peut poser problème pour calculer naïvement une variation temporelle comme illustré figure 4.4. De manière générale, pour chaque trajectoire, on observe un plot toutes les 1.8 secondes en moyenne. Cette moyenne résulte d'un écart entre deux plots de deux secondes dans 80% des cas et d'une seconde pour les 20% restant.

À chaque plot radar Mode-C, on associe un vent et une température en utilisant une grille météorologique (cf. section 4.1.1). De même, à chaque vol, on associe des informations issues des plans de vols (cf. section 4.1.1) telles que le niveau de vol et la vitesse de croisière souhaités.

```
<?xml version="1.0"?>
<Message><Header><DateTime>2006-07-01T00 :02 :00Z</DateTime><Source Site="CRNA
    " Module="STR">STR</Source><Target Site="CRNA" Module="BOSSA_SERVER">
    BOSSA_SERVER_CRNA</Target><CDMVersion>V1.0.9</CDMVersion><MessageType>
    Plots</MessageType><MessageInfo>STR_CRNAN</MessageInfo></Header><Plot><
    PlotId><CallSign>WLX889</CallSign><Time>00 :04 :31Z</Time><TrackNumber>527<
    /TrackNumber></PlotId><Position><X>139765.80</X><Y>-346736.81</Y><Z>
    190.00</Z></Position><Velocity><Vx>46</Vx><Vy>-120</Vy><Vz>0</Vz></
    Velocity><PlotData><SSR>3371</SSR><VehicleFleetId>0</VehicleFleetId><
    Acrft>ATP</Acrft><Dep>LFPB</Dep><Arr>LFML</Arr></PlotData></Plot></
    Message>
```

FIGURE 4.2 – Extrait d'un fichier de plots radar Mode-C. Chaque plot contient l'heure (00 :04 :31 Z), l'identifiant (WLX889), le type avion (ATP), l'aéroport de départ (LFPB Paris-Le Bourget) et d'arrivée (LFML Marseille), la position (X = 139 765,80 m, Y = −346 736,81 m, Z = 19 000 ft) et la vitesse (V_X = 46 kts, V_Y = −120 kts, V_Z = 0) de l'avion.

Trajectoires Mode-S

Les données de trajectoires Mode-S sont issues d'un radar Mode-S EHS situé à la Direction de la Technique et de l'Innovation (DTI) de Toulouse. On dispose de 245 jours d'enregistrements répartis de février 2011 à décembre 2012. Pour chaque trajectoire, on observe en moyenne un plot toutes les 4 secondes, avec parfois des « trous » pour lesquels une série de plots est manquante comme illustré figure 4.5. Ces « trous » limitent grandement le nombre de trajectoires exploitables. Ces « trous » sont dû à une limitation du débit de la liaison entre le radar et le site de l'ÉNAC. La collecte et l'archivage des données à des fin de recherche est au stade artisanal en France. Comme pour les trajectoires Mode-C, chaque plot est daté et contient la position et la vitesse de l'avion. À ces informations viennent s'ajouter d'autres grandeurs issues des équipements à bord de l'avion : l'inclinaison, le cap magnétique, le taux de virage, la vitesse verticale inertielle, la vitesse verticale barométrique, la vitesse sol, la True AirSpeed (TAS), l'Indicated AirSpeed (IAS) et le Mach. Dans ce qui suit, on considère l'IAS égale à la CAS.

On a accès à plusieurs grandeurs représentant chacune une vitesse. Les relations les liant entre elles ne dépendent que des conditions météorologiques ; ces relations vont donc nous permettre de déduire des grandeurs météorologiques telles le vent et la température. Le vent subi par l'avion correspond à la différence entre la TAS et la vitesse sol. De même, la température est proportionnelle au carré du quotient de la TAS et du Mach.

4.1.3 Lissage des données

Comme vu dans la section 4.1.2, on dispose de deux jeux de données. De ces jeux de données, on souhaite extraire deux jeux de trajectoires à exploiter. Des informations telles l'énergie et surtout sa variation nous semblent nécessaires. En effet, dans les modèles physiques dit à *énergie totale*, tel BADA par exemple, la modélisation des forces projetées

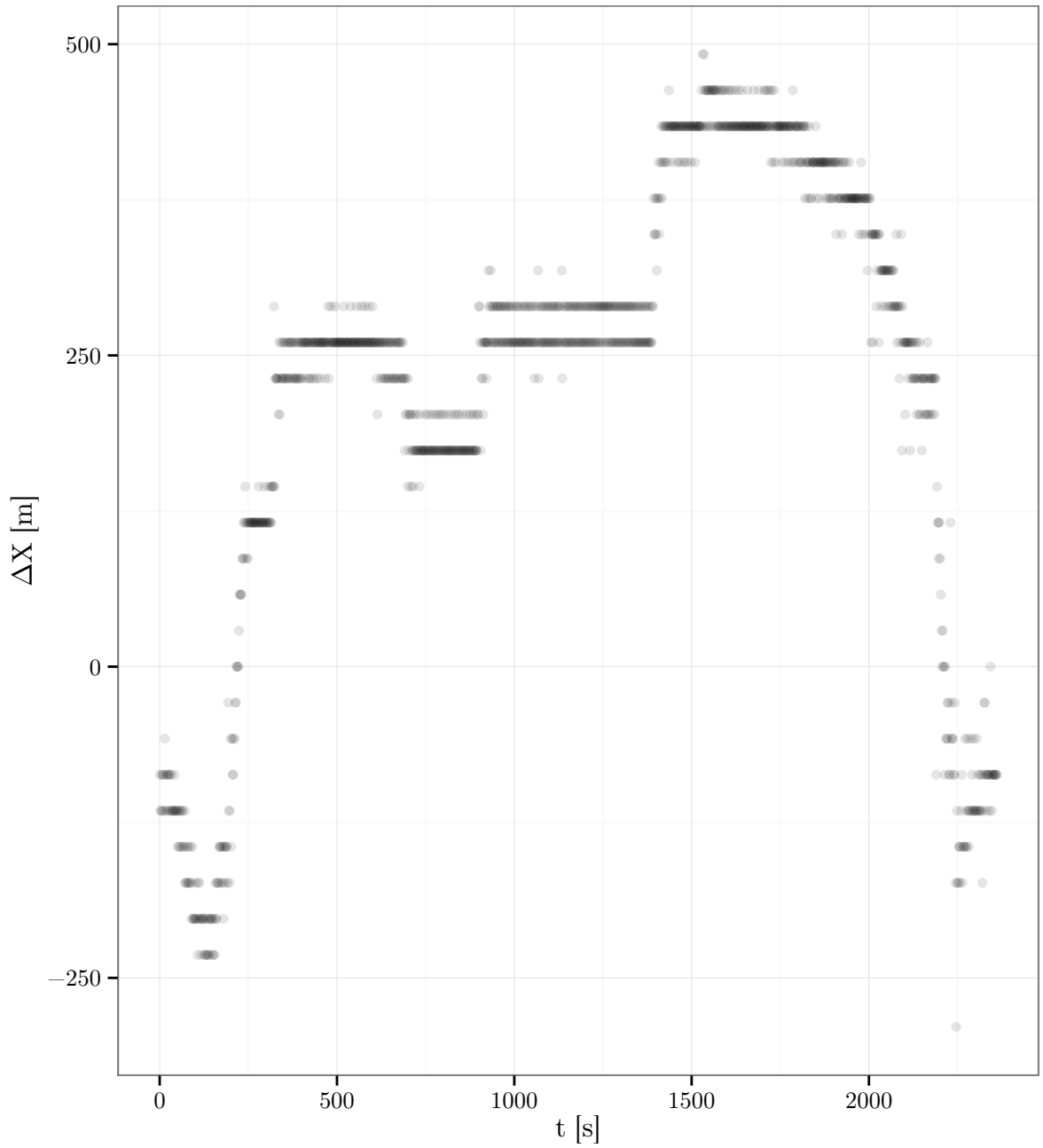


FIGURE 4.3 – Pour une trajectoire donnée issue du Mode-C, ce graphique représente la variation ΔX de la coordonnée X entre deux plots consécutifs en fonction du temps. On observe une quantification sur X ; celle-ci est présente pour toutes les grandeurs et pour toutes les trajectoires d'avions. Pour nos données, la granularité de la quantification sur X et Y est de 29 m.

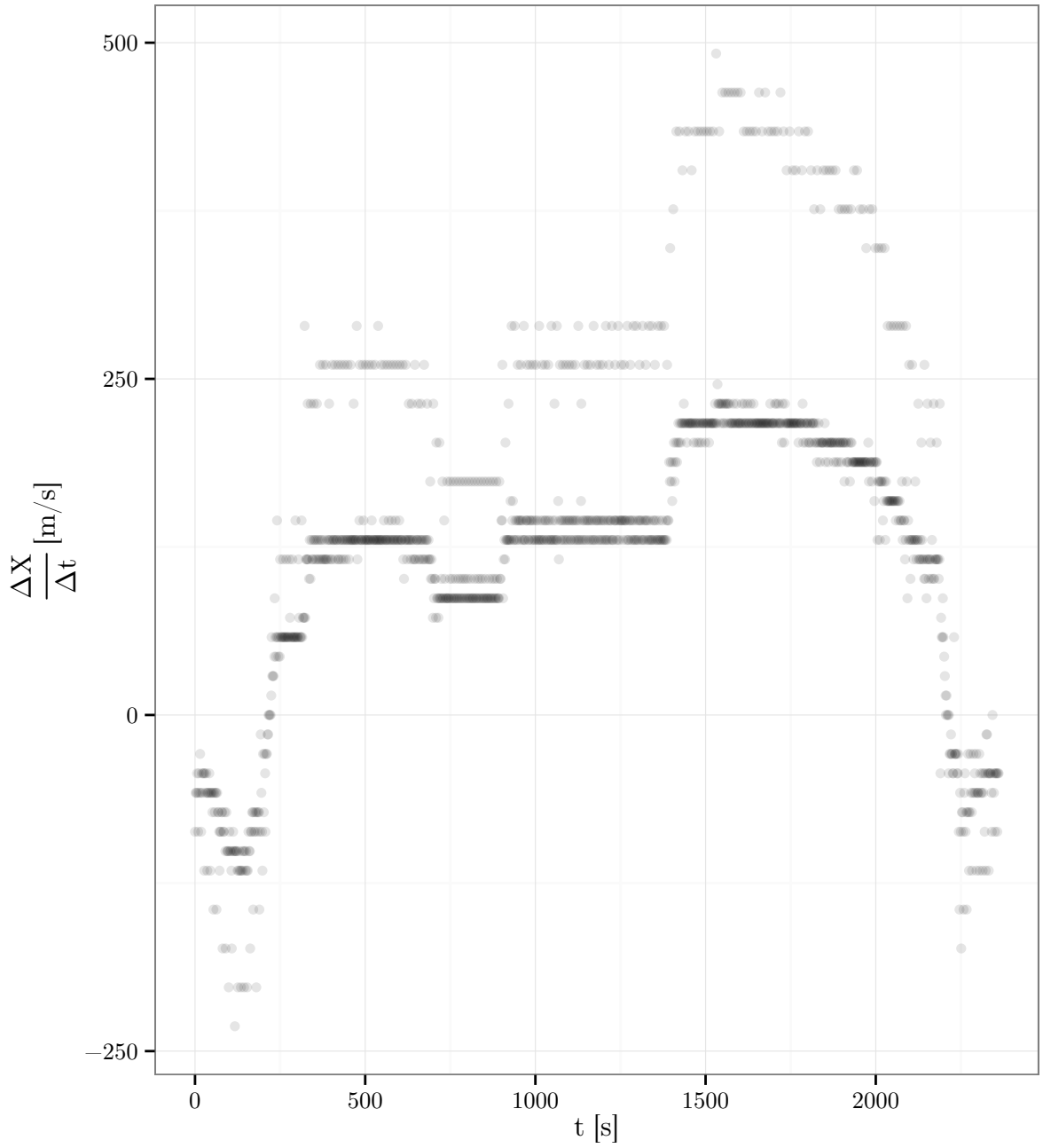


FIGURE 4.4 – Pour une trajectoire donnée issue du Mode-C, ce graphique représente les différents taux d'accroissements de X calculés entre des plots successifs. On observe la superposition de deux courbes, l'une valant le double de l'autre. Celle qui vaut le double correspond à des Δt d'une seconde. Cette superposition de deux courbes qui n'a pas lieu d'être est une conséquence de la datation des plots à la seconde près.

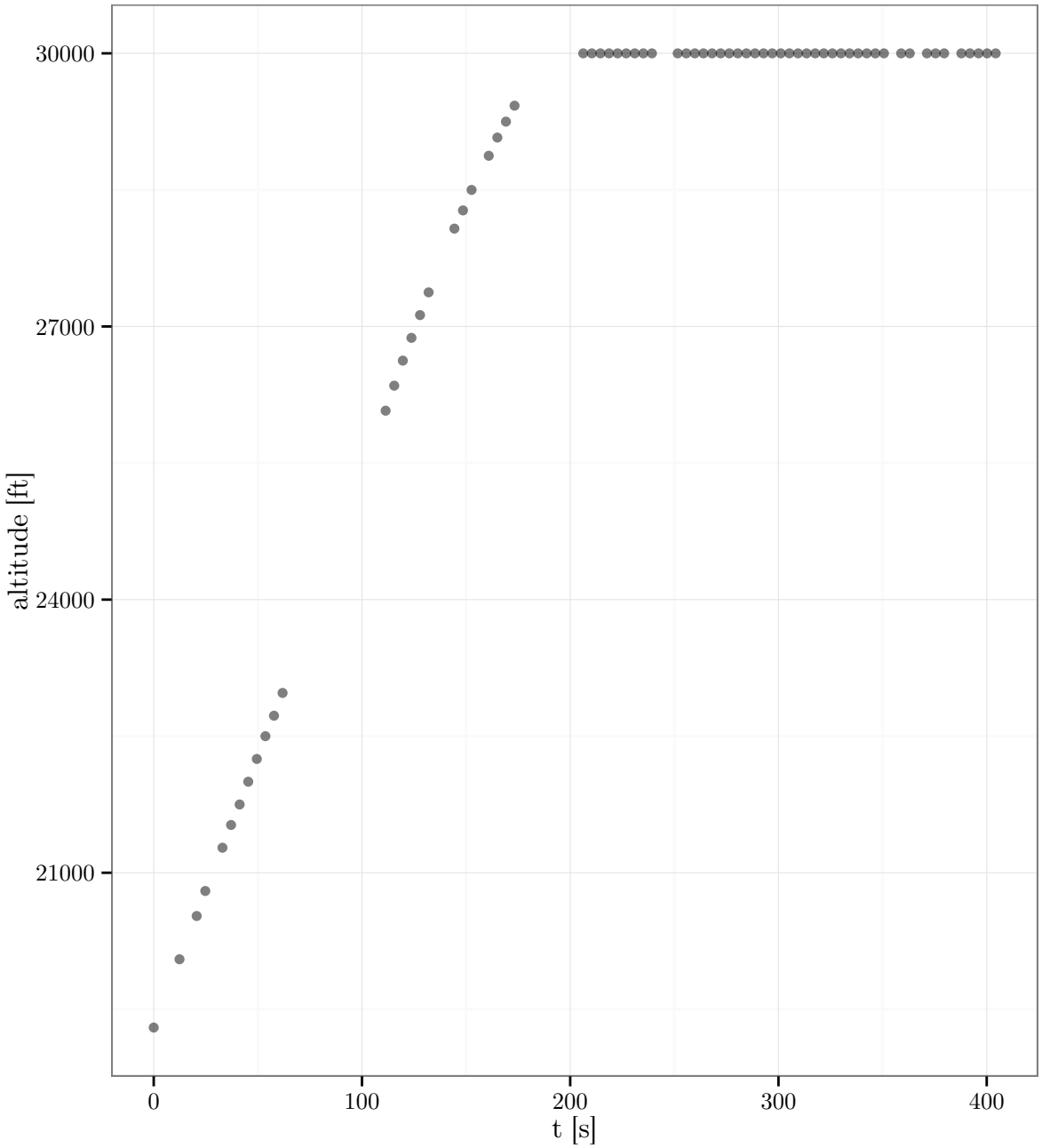


FIGURE 4.5 – Pour une trajectoire donnée issue du Mode-S, ce graphique représente l'altitude en fonction du temps. On observe des « trous » dans la trajectoire.

sur le vecteur vitesse ne sert qu'à estimer cette variation d'énergie. Partant de cette nécessité, il nous faut obtenir un taux de montée et une accélération. L'obtention de ces informations requiert un lissage. Prendre naïvement le taux d'accroissement n'est pas satisfaisant comme illustré figure 4.4.

À partir d'une succession de n points à lisser, $(x_i, y_i) \in \mathbb{R}^2$ avec $x_1 \leqslant \ldots \leqslant x_n$, le lissage permet de n'en retenir que la tendance générale en éludant des variations de faible amplitude ou de fréquence élevée. Dans la plupart des méthodes de lissage, l'utilisateur ajuste les paramètres de la méthode pour sélectionner une tendance générale qui l'intéresse. Si les variations enlevées provenaient d'un bruit, le lissage va permettre de réduire le bruit des données.

Concernant le lissage de trajectoires radar d'avions, différents algorithmes de lissage ont été expérimentés tels les splines [Mehadhebi 99b], les filtres de Kalman [Mehadhebi 99a] et les B-splines à rigidité variable [Mehadhebi 02]. On a choisi d'utiliser un lissage par spline. Une spline d'ordre m est une fonction $C^{2(m-1)}$ définie par morceaux à l'aide de polynômes de degré $2m-1$. Les points séparant ces polynômes sont appelés *nœuds*. Les splines de lissage sont plus lisses que les splines d'interpolation car on les autorise à ne pas passer par les points (x_i, y_i). De cet ajout de liberté résulte un compromis à résoudre par l'utilisateur entre le caractère lisse de la spline et l'approximation faite des points (x_i, y_i). Ce compromis est traduit par le critère $L(f, \lambda)$ à minimiser, défini en 4.1, qui correspond à une pondération de deux termes, chacun de ces termes représentant un des deux critères antagonistes : la somme des erreurs au carré permet de prendre en compte l'erreur d'approximation ; et la norme L^2 de la dérivée d'ordre m sur l'intervalle $[x_1; x_n]$ représente le caractère lisse de la spline. Le caractère lisse peut être défini autrement en modifiant la pénalité ; en considérant la norme L^2 d'une combinaison linéaire des dérivées de f, [Ramsay 97] spécifie ainsi une pénalité qui va considérer comme lisse des oscillations d'une certaine fréquence. La détermination de la spline f minimisant $L(f, \lambda)$ est faite avec une complexité temporelle en $O(n)$ ([Reinsch 67]).

$$L(f, \lambda) = \frac{1}{n}\sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int_{x_1}^{x_n} f^{(m)}(t)^2 dt \tag{4.1}$$

Le paramètre de lissage λ doit être choisi. La validation-croisée ([Stone 74]), présentée plus en détail section 3.1.2, peut être utilisée pour sélectionner ce paramètre. Dans la pratique, cette méthode est décevante comme l'illustre la figure 4.6. Le paramètre λ semble sous-estimé ; la trajectoire lissée exhibe des sauts peu probables pour une trajectoire d'avion. Cette sous-estimation du paramètre de lissage λ est aussi observée dans la thèse de Simpkin sur l'estimation des dérivées par le biais d'un lissage ([Simpkin 10]). Pour contourner ce problème, la fonction *D1D2* du package R *sfsmisc* servant à estimer les dérivées premières et secondes rajoute un terme constant arbitraire au λ estimé par validation croisée.

Pour notre problème, on utilise un critère plus pragmatique pour sélectionner le paramètre de lissage λ. Une fois lissée, la trajectoire sera échantillonnée avec un certain pas de temps Δt. On souhaite que la trajectoire échantillonnée ainsi que ses dérivées apparaissent co-

hérentes. Par exemple, si on a calculé un taux de montée sur le point courant, on s'attend à ce que le point suivant ait une altitude cohérente avec ce taux de montée. On veut aussi que la trajectoire échantillonnée soit proche des points effectivement mesurés. On a choisi de trancher ce compromis entre la cohérence de la trajectoire échantillonnée et la précision en minimisant l'écart entre les observations et une trajectoire reconstituée. Cette trajectoire reconstituée est construite à partir des dérivées calculées tous les Δt que l'on intègre numériquement. Fixer un Δt va donc nous permettre, indirectement, de sélectionner un λ. Le Δt sera le même pour toutes les trajectoires mais le λ correspondant sera ajusté pour chaque trajectoire.

Il nous faut également choisir l'ordre de la spline m. Si l'on souhaite estimer la k^{ime} dérivée, les discussions issues de [Silverman 85] suggèrent de prendre $m \geqslant k+2$. Ainsi, si on veut uniquement la dérivée, on prendra $m \geqslant 3$ et si l'on veut en plus l'accélération, on prendra $m \geqslant 4$. Sous certaines hypothèses détaillées dans [Ragozin 83], les dérivées estimées convergent vers les vraies dérivées au sens L^2.

Une fois le lissage choisi, les dérivés pouvant nous intéresser sont calculées de proche en proche en utilisant la formule :

$$\frac{d}{dt}\left[f(x_1(t),\ldots,x_n(t))\right] = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_1(t),\ldots,x_n(t))\frac{\partial x_i}{\partial t}(t)$$

Pour le Mode-C par exemple, on a lissé x, y et H_p. On a donc accès à leur dérivées également. La grille météo nous permet de décrire le vent w en fonction de la position de l'avion (x, y, H_p) et de la date t. Ainsi, en appliquant la formule précédente à $w(x(t), y(t), H_p(t), t)$, on obtient la dérivée du vent w par rapport au temps. Cette dérivée va pouvoir être à son tour utilisée pour calculer la dérivée d'une expression faisant intervenir le vent, comme la TAS par exemple.

Paramètres de lissage pour le Mode-C

Pour choisir un lissage, on va regarder l'erreur faite entre la trajectoire lissée et la trajectoire brute. Cette erreur va dépendre des paramètres de lissage choisis. Le tableau 4.1 présente les erreurs obtenues suivant le lissage. Les différentes Root Mean Square Error (RMSE) obtenues sont du même ordre de grandeur. Toutefois, l'erreur maximum obtenue est bien plus faible pour la spline d'ordre 3. Comme illustré figure 4.7, on peut observer des écarts importants entre les valeurs lissées, en particulier sur les bords. Concernant cet exemple, la spline d'ordre 3 est la plus proche des données brutes, en particulier sur les bords. Ce constat se répétant sur plusieurs trajectoires nous conduit à choisir une spline d'ordre 3 pour les trajectoires Mode-C. Pour λ, on a choisi $\Delta t = 15$ s car les erreurs sont assez similaire entre $\Delta t = 10$ s et $\Delta t = 15$ s et l'on préfère choisir les trajectoires les plus lisses pour des erreurs du même ordre de grandeur. Toutefois, on a écarté $\Delta t = 20$ s à cause des erreurs maximum plus élevées.

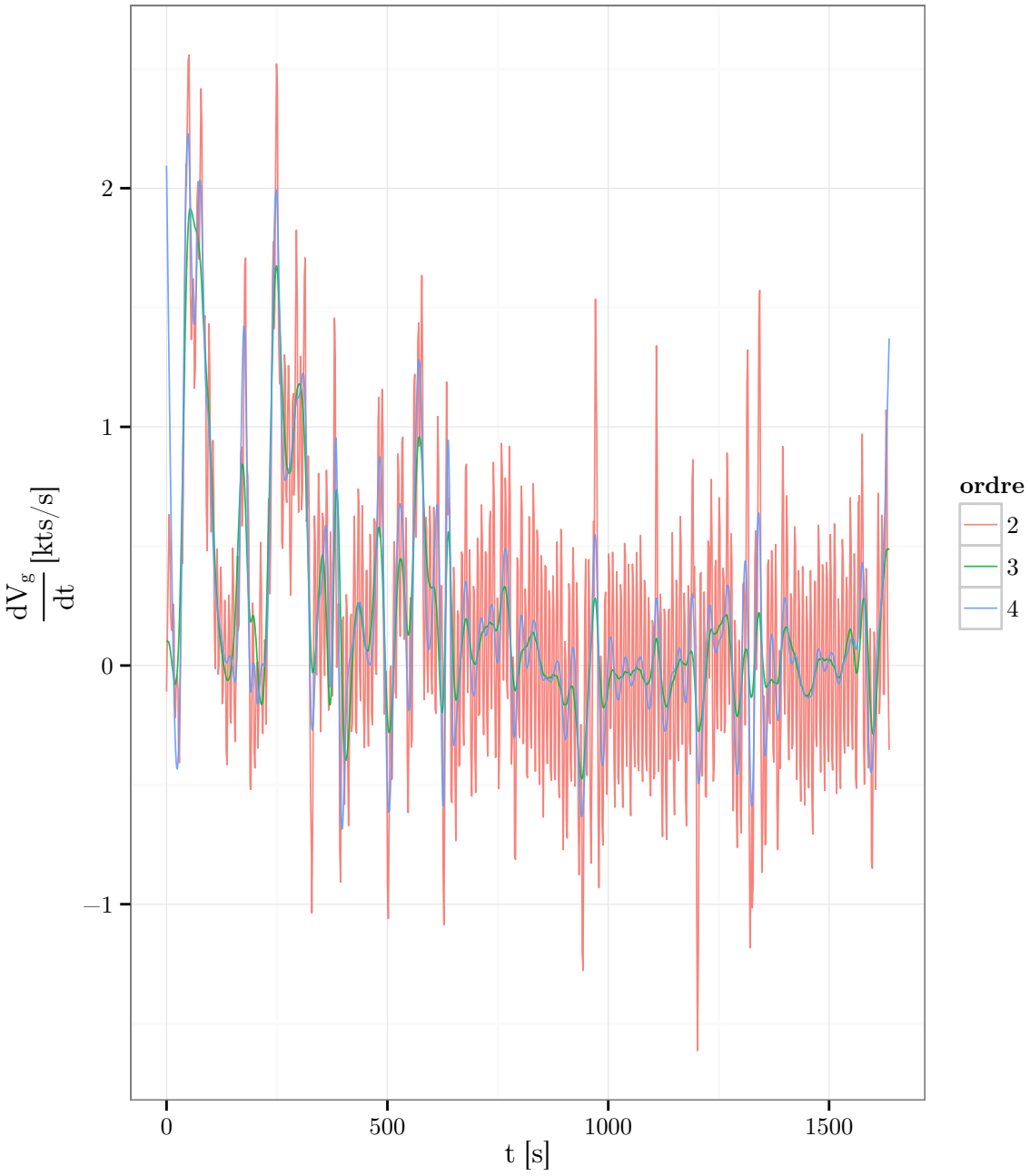


FIGURE 4.6 – Pour un avion donné, ce graphique représente l'estimation de l'accélération par rapport au sol par une spline dont le paramètre de lissage λ a été ajusté par validation croisée. On observe des oscillations rapides de l'accélération, en particulier pour la spline d'ordre 2.

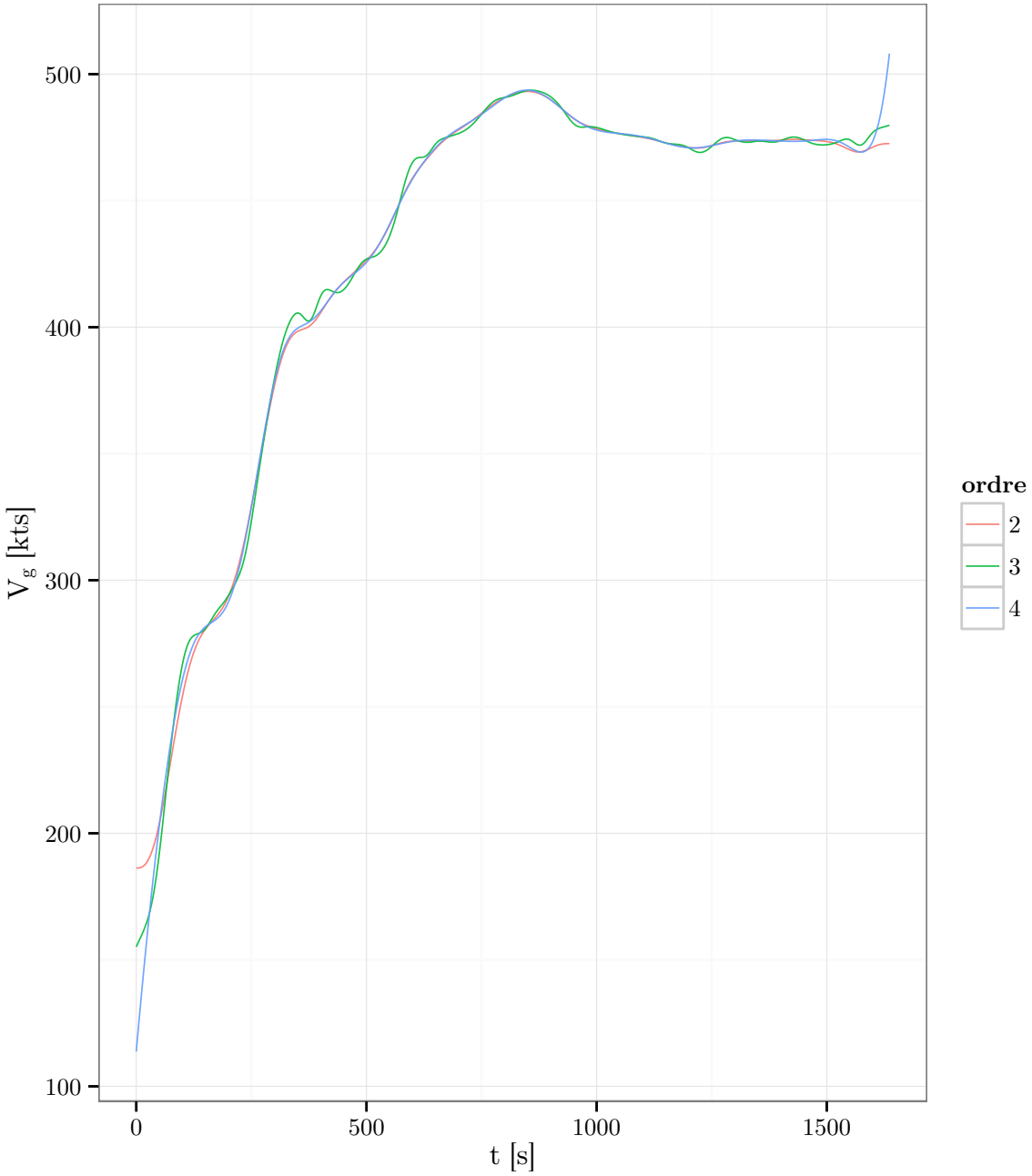


FIGURE 4.7 – Pour un avion donné, ce graphique représente l'estimation de la vitesse par rapport au sol par des splines d'ordres différents. Le paramètre de lissage λ a été ajusté avec $\Delta t = 15s$.

TABLE 4.1 – Statistiques sur l'écart entre les trajectoires lissées et les plots observés pour les trajectoires Mode-C.

λ	ordre	x [m]		y [m]		H_p [ft]	
		rmse	max	rmse	max	rmse	max
$\Delta t = 10$	2	77.22	1749.12	79.97	1828.93	36.16	371.92
$\Delta t = 10$	3	42.40	433.13	51.14	448.68	36.10	344.15
$\Delta t = 10$	4	55.56	1807.99	60.68	2242.47	36.81	348.75
$\Delta t = 15$	2	117.27	1892.43	110.12	1833.88	43.24	395.90
$\Delta t = 15$	3	49.01	468.77	54.77	543.79	38.22	344.15
$\Delta t = 15$	4	80.03	2671.71	79.23	2446.30	38.69	348.75
$\Delta t = 20$	2	153.45	1892.43	135.28	1833.88	46.18	973.67
$\Delta t = 20$	3	50.23	732.01	55.42	863.39	39.50	344.15
$\Delta t = 20$	4	121.16	2701.72	104.60	2445.53	41.06	521.07

Paramètres de lissage pour le Mode-S

Pour les trajectoires Mode-S, le tableau 4.2 présente les erreurs obtenues suivant le lissage. Les erreurs sont du même ordre de grandeur pour tous les paramétrages. Pour les même raisons que pour les trajectoires Mode-C, on préfère la spline d'ordre 3. Pour des raisons de simplification, les trajectoires Mode-C et Mode-S seront échantillonnées avec le même pas de temps 15 s.

TABLE 4.2 – Statistiques sur l'écart entre les trajectoires lissées et les plots observés pour les trajectoires Mode-S.

λ	ordre	cap [°]		route [°]		V_{sol} [kts]		CAS [kts]		TAS [kts]		Mach[-]		H_p [ft]	
		rmse	max	rmse	max	rmse	max	rmse	max	rmse	max	rmse	max	rmse	max
$\Delta t = 10$	2	0.19	5.81	0.15	5.46	0.52	4.72	0.49	7.23	0.69	8.11	0.00	0.01	7.73	104.19
$\Delta t = 10$	3	0.14	6.84	0.11	5.27	0.48	5.54	0.43	6.09	0.62	6.60	0.00	0.01	6.04	87.18
$\Delta t = 10$	4	0.15	7.33	0.13	7.49	0.50	5.90	0.45	9.12	0.65	8.16	0.00	0.01	6.31	156.96
$\Delta t = 15$	2	0.21	5.81	0.19	5.46	0.54	10.54	0.51	8.75	0.74	8.11	0.00	0.01	7.77	181.80
$\Delta t = 15$	3	0.20	10.91	0.17	10.99	0.55	7.76	0.53	11.94	0.74	13.12	0.00	0.02	7.36	141.80
$\Delta t = 15$	4	0.23	9.60	0.19	11.41	0.59	8.89	0.59	11.81	0.79	12.52	0.00	0.02	8.44	164.04
$\Delta t = 20$	2	0.29	14.77	0.24	10.09	0.70	9.08	0.74	13.53	1.01	14.41	0.00	0.02	8.76	183.72
$\Delta t = 20$	3	0.31	14.10	0.28	14.56	0.67	9.12	0.68	11.94	0.91	13.12	0.00	0.02	10.56	219.99
$\Delta t = 20$	4	0.32	18.81	0.29	19.44	0.71	13.03	0.73	11.81	0.94	16.50	0.00	0.02	10.94	200.45

4.1.4 Filtrage des trajectoires

On s'intéresse à la prédiction de la trajectoire d'un avion en phase de montée. Il nous faut donc isoler les trajectoires d'avions en montée. Pour ce faire, on va utiliser le taux de montée que l'on a estimé. Comme celui-ci a été lissé, on va exiger que le taux soit supérieur à un certain seuil strictement positif $ROCD_{seuil}$ pour éviter de prendre une courte phase de palier à l'intérieur d'un segment de montée. De plus, on veut éviter de prendre les phases de transition entre une phase de montée et de palier. On va donc retirer une partie de la trajectoire en début et fin de montée. On a choisi de retirer 40 secondes en début et fin de montée. Le seuil $ROCD_{seuil}$ a été fixé à 150 ft/min après essai de différentes valeurs.

4.1.5 Construction des exemples

Comme vu dans la section 4.1.3, le paramètre de lissage λ a été choisi pour que la trajectoire échantillonnée tous les Δt et les dérivées associées soient cohérentes. Le paramètre de lissage Δt donnant satisfaction est $\Delta t = 15s$, il est donc naturel d'échantillonner avec un pas de temps de $15s$ pour lequel un critère de cohérence a été optimisé.

Une fois que l'on a échantillonné les montées que l'on a filtrées dans la section 4.1.4, il faut en extraire un jeu d'exemples du problème qui nous intéresse. Le problème traité ici est la prédiction de trajectoires. À partir d'une trajectoire donnée, on va former un ou plusieurs exemples. Un exemple est constitué à partir de $p + q$ points consécutifs de la trajectoire. Les p premiers points forment la trajectoire passée et les q points suivants constituent la trajectoire future. L'exemple ainsi formé correspond bien à un problème de prédiction pour lequel on souhaite prédire q points futurs connaissant les p points passés.

On a choisi un pas de temps de $15s$ pour échantillonner nos trajectoires. Il est naturel de commencer à échantillonner à partir du début de la trajectoire. Toutefois, il peut être intéressant de décaler la date à partir de laquelle on échantillonne la trajectoire. Par exemple, si l'on considère un avion à une altitude donnée H_{p_0}, il est intéressant d'avoir un jeu d'exemples pour lequel le $p^{ième}$ point se situe à l'altitude H_{p_0}. On va ainsi chercher l'instant t_0 auquel l'avion passe H_{p_0} et échantillonner $p-1$ points dans le passé et q points dans le futur à partir de la date t_0. Pour trouver une telle date t_0, on va procéder par une recherche dichotomique sur la fonction $H_p(t)$ issue du lissage. En effet, l'altitude peut raisonnablement être considérée comme croissante avec le temps puisqu'on a sélectionné les trajectoires avec un taux de montée supérieur à $ROCD_{seuil}$. La figure 4.8 illustre un exemple obtenu avec $H_{p_0} = 18\ 000$ ft, $p = 11$ et $q = 40$.

4.2 Analyse des exemples

Dans cette section, on analyse les exemples issus des trajectoires radar au regard de la modélisation des montées qui est faite dans BADA. Les performances en montée diffèrent suivant le type avion. Pour cette analyse, seules les trajectoires d'Airbus A320, le type avion le plus représenté dans nos données, sont étudiées. Il nous faut maintenant choisir une altitude H_{p_0}. À partir de l'altitude H_{p_0} et de $p = 11$ points de la trajectoire passée, on

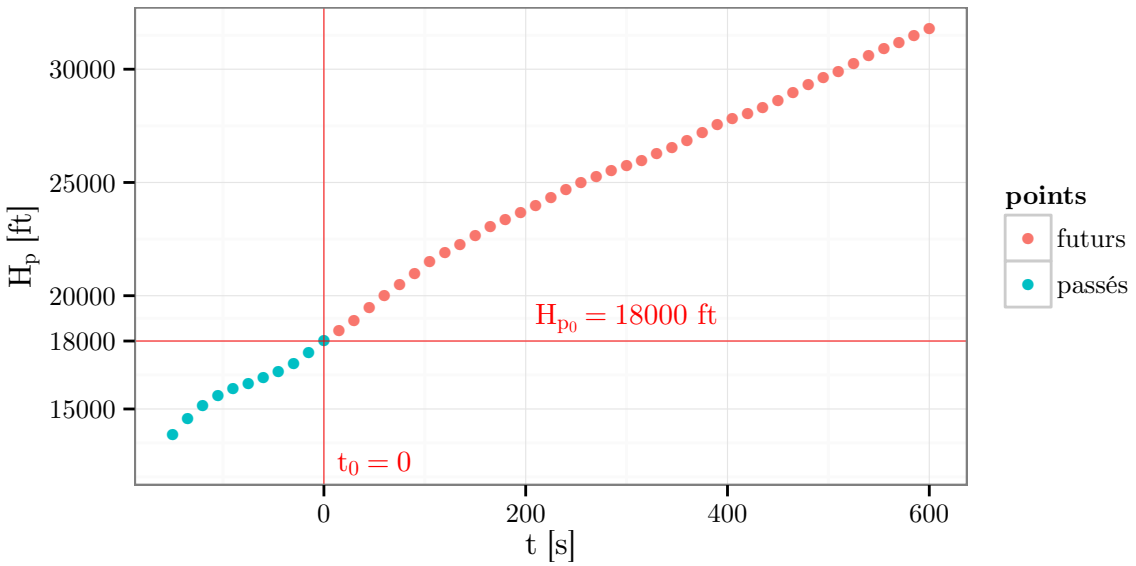


FIGURE 4.8 – Cette figure représente une trajectoire de notre jeu d'exemples.

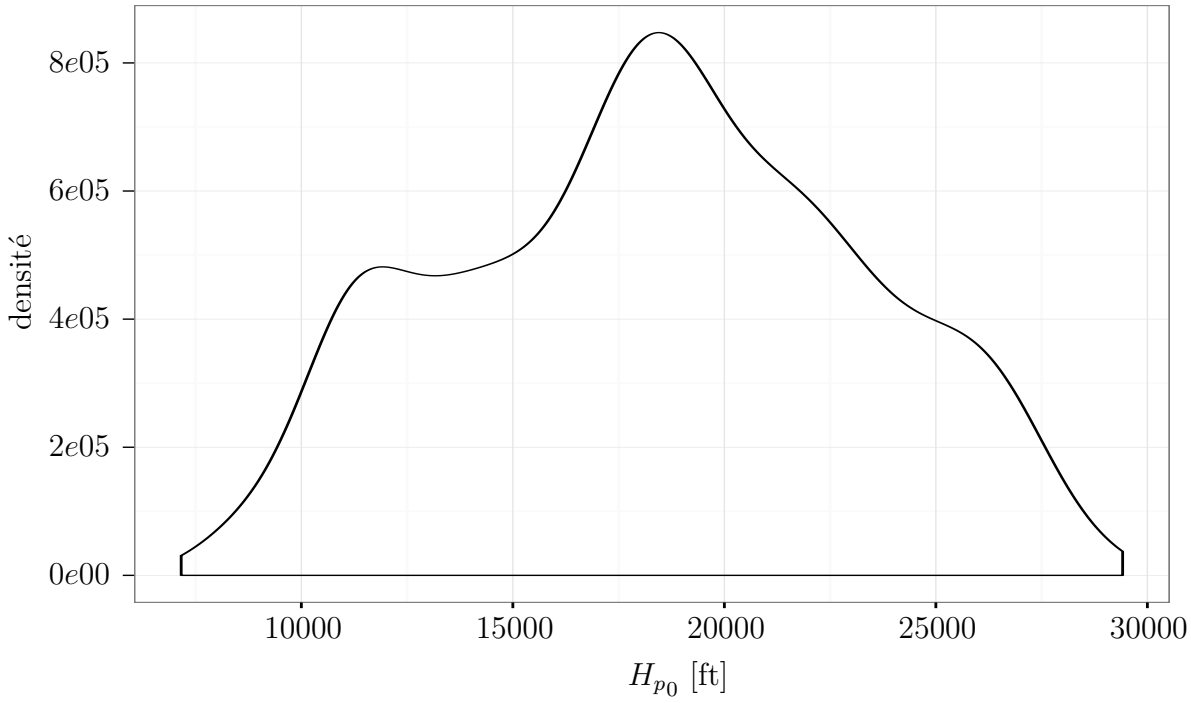


FIGURE 4.9 – Pour les données Mode-S, on a considéré tout les H_{p_0} possibles pour chaque trajectoire, puis on a tracé la densité de ces H_{p_0} possibles.

souhaite prédire l'altitude avec un horizon temporel de 10 minutes ce qui conduit à choisir $q = 40$. Cet horizon temporel est le minimum requis pour que la prédiction puisse être utile à la déconfliction de trajectoires en phase tactique. Selon la figure 4.9, le choix $H_{p_0} = 18000ft$ est celui qui maximise le nombre d'exemples extrait à partir de nos trajectoires Mode-S.

À la fin, on obtient deux jeux d'exemples. On a extrait 4 939 exemples des données Mode-C et 1 062 exemples des données Mode-S.

4.2.1 Profils de température observés

L'atmosphère ISA modélise la variation de la température en fonction de l'altitude H_p. Au niveau de la mer, on a un écart ΔT entre la température mesurée et modélisée [1] par l'atmosphère ISA. L'équation différentielle $dT = \beta_T dH_p$ (cf. section 2.2) est une hypothèse de l'atmosphère ISA, avec β_T constante par plages d'altitude. D'après cette équation, l'écart ΔT est conservé à des altitudes H_p plus hautes. Ainsi à partir de chaque point d'altitude H_p, on peut calculer l'écart ΔT en faisant la différence entre la température mesurée et la température modélisée par l'atmosphère ISA à l'altitude H_p. Les figures 4.10 et 4.11 représentent cet écart en fonction de l'altitude pour nos jeux de trajectoires échantillonnées Mode-C et Mode-S. Avec ces figures, on remarque que les ΔT mesurés ne sont pas constants au cours de la montée. Ceci ne correspond pas à la modélisation de la température faite dans l'atmosphère ISA. Du fait que notre grille météo s'arrête à 34000 ft, les températures et vents au-dessus de 34000 ft proviennent uniquement du dernier niveau de la grille. Au-dessus de 34000 ft, ΔT est supposé constant.

L'équation $dT = \beta_T dH_p$ ne se vérifie donc pas sur les données. Cet écart avec le modèle d'atmosphère n'a pas de conséquence sur la validité des équations du modèle BADA. Les équations utilisées sont celles qui sont valables localement. Le profil de la température mesurée n'est exploité que lorsque l'on souhaite calculer l'altitude géopotentielle H ou l'altitude géodésique h. En effet, l'obtention de ces valeurs nécessite le calcul d'une intégrale faisant intervenir la température mesurée. Ce calcul n'a pas été nécessaire dans cette étude.

4.2.2 Profils de vitesse observés

Le profil de vitesse est une donnée très importante pour prédire l'altitude future. En effet, ce profil va contraindre le partage de la puissance entre l'énergie cinétique et potentielle.

Dans le modèle BADA, comme explicité section 2.5.2, les profils de vitesse sont modélisés au travers du couple (CAS_{cible},$Mach_{cible}$). Pour l'Airbus A320, BADA 3.9 modélise la montée avec le couple ($CAS_{cible} = 310$ kts,$Mach_{cible} = 0.78$). Avec un profil de ce type, une fois le couple (CAS_{cible},$Mach_{cible}$) choisi, la vitesse est une fonction de l'altitude. Pour confronter cette modélisation aux profils observés, on a tracé la vitesse CAS en fonction de l'altitude H_p sur les figures 4.12 et 4.13. Sur ces figures, on remarque bien l'accélération passant 10000 ft qui est induite par la limitation à 250 kts en dessous du FL100. Sur la

1. $15°C$ au niveau de la mer.

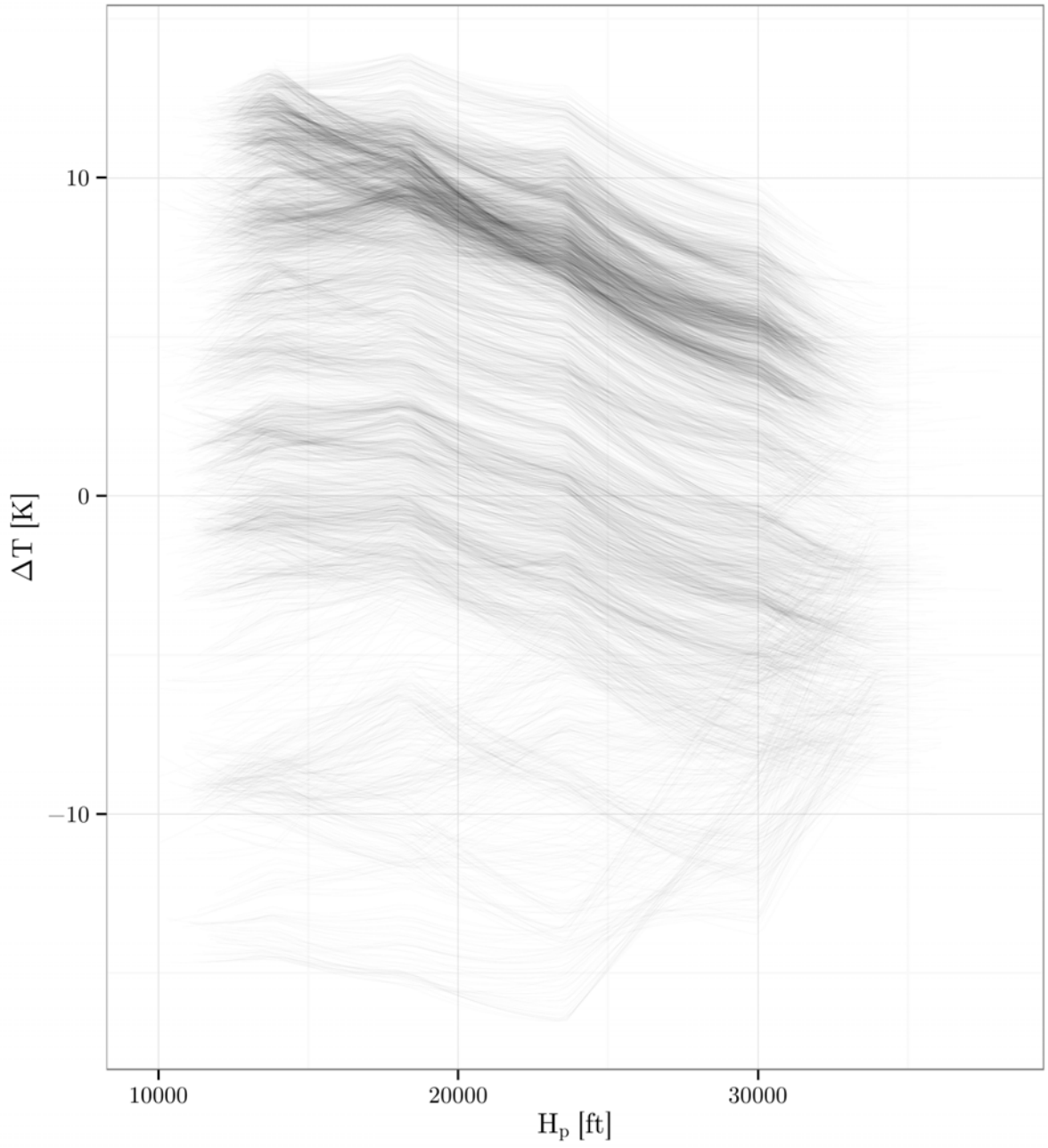


FIGURE 4.10 – Ce graphique représente l'écart ΔT en fonction de l'altitude H_p, pour les trajectoires Mode-C.

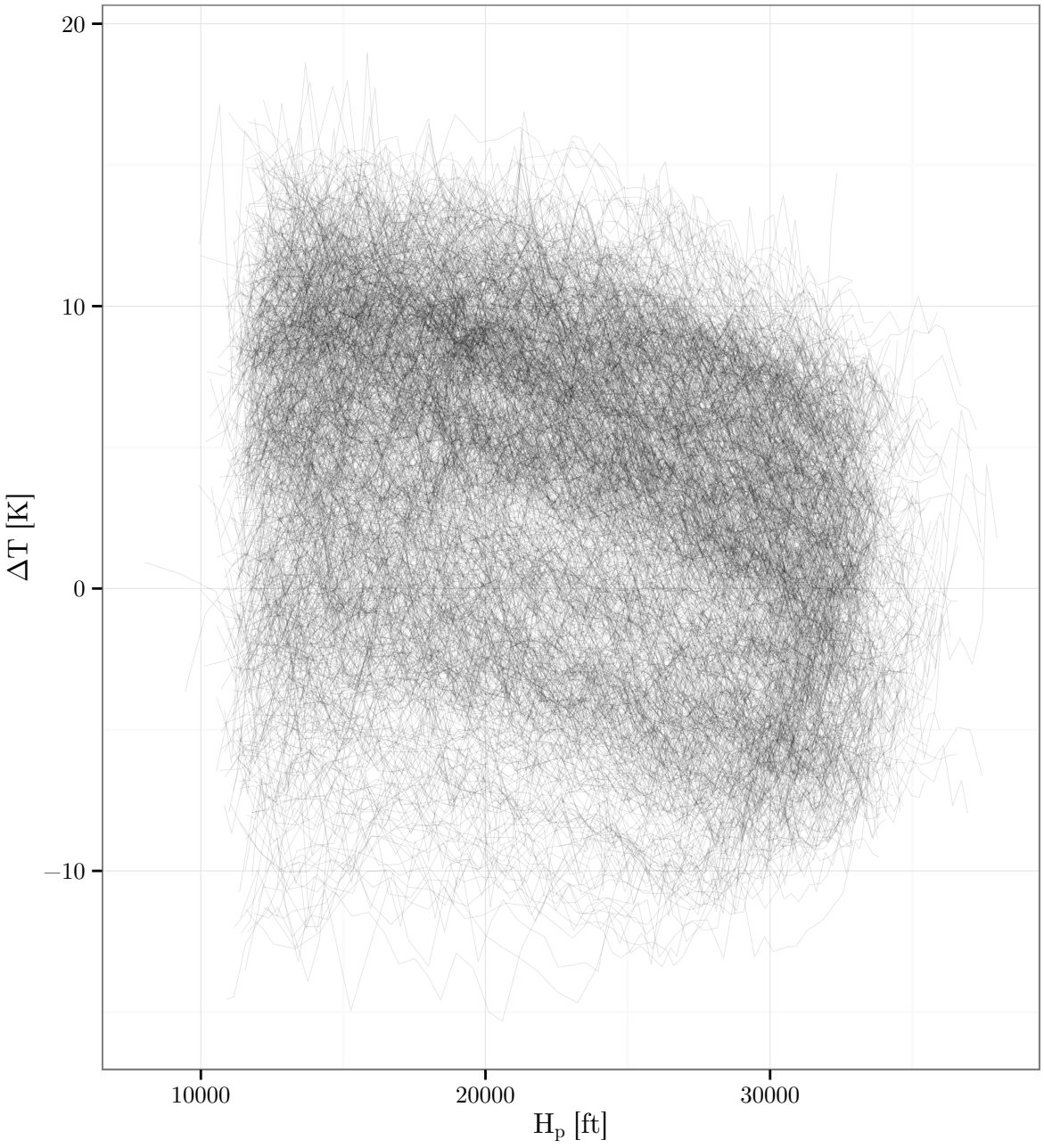


FIGURE 4.11 – Ce graphique représente l'écart ΔT en fonction de l'altitude H_p, pour les trajectoires Mode-S.

figure 4.12, on voit apparaître trois accumulations, une aux alentours de 280 kts qui perd en intensité avec l'altitude suggérant que les avions concernés ont accéléré. Les deux autres correspondent à des avions stables sur un profil de type (CAS_{cible},$Mach_{cible}$). On a une accumulation à 305 kts et l'autre à 340 kts qui rejoignent une accumulation selon une droite décroissante qui correspond aux segments à Mach constant.

Pour chaque trajectoire, on a ajusté un profil (CAS_{cible},$Mach_{cible}$) au profil de vitesse observé aux points d'altitude supérieur à H_{p0}. La méthode d'ajustement est décrite en annexe B. Avec la figure 4.14, on remarque deux accumulations, une à (CAS_{cible} = 305 kts,$Mach_{cible}$ = 0.78) et une autre plus resserrée à (CAS_{cible} = 340 kts,$Mach_{cible}$ = 0.80). Cette dernière contient 33% des avions au départ de LFPO et 18% de ceux au départ de LFPG. L'erreur d'ajustement est bien plus faible pour les trajectoires Mode-S que les trajectoires Mode-C. La médiane de la RMSE entre la la TAS observée et la TAS ajustée est de 1.30 kts pour les données Mode-S et de 4.45 kts pour les données Mode-C.

Entre ces deux accumulations sur les figures 4.12 et 4.13, on distingue des avions qui ont une phase d'accélération. Pour mieux analyser ceci, les figures 4.16 et 4.17 présentent l'accélération en fonction de la vitesse pour les trajectoires ne correspondant pas à un profil type (CAS_{cible},$Mach_{cible}$). Seules les trajectoires dont l'erreur sur la TAS après ajustement d'un profil (CAS_{cible},$Mach_{cible}$) fait partie des 25% les plus élevées sont considérées. On a uniquement affiché les points d'altitudes supérieurs à H_{p0} pour voir les changements de vitesses au cours de la montée à prédire. Ces figures permettent de voir que les phases d'accélération ne sont pas forcément prépondérantes. Plus de la moitié des avions ont une accélération $\frac{dCAS}{dt}$ inférieure à -0.5 kts/s. Les avions décélèrent plus fortement qu'ils n'accélèrent. Parmi ces avions qui décélèrent, il y a deux comportements type : les avions qui accélèrent tout de suite après une phase de décélération qui arrive tôt dans la montée ; et ceux qui décélèrent tardivement et plus violemment que ce que requiert le suivi d'un segment Mach constant. Ces décélérations sont accompagnées d'un taux de montée élevé. Il y a tout de même des avions en vitesse stabilisée à CAS constante qui accélèrent pour de nouveau rejoindre une phase stabilisée avec une CAS plus élevée. Il y a également des avions pour lesquels la première phase constante n'apparaît pas, ils sont en phase d'accélération dés H_{p0} ; on peut supposer que l'avion a commencé à accélérer avant H_{p0} pour rejoindre sa deuxième phase à CAS constante.

4.2.3 Étude de la variation d'énergie spécifique

La variation d'énergie spécifique (SER) $g_0\frac{dH}{dt} + TAS\frac{dTAS}{dt}$ est une grandeur importante dans la prédiction de trajectoire. Dans un modèle à énergie totale, la modélisation des forces permet le calcul de cette grandeur. En effet, La modélisation des forces permet le calcul de l'excès de puissance spécifique (SEP) qui est égal par définition à $\frac{Thr-D}{m}TAS - \frac{d\overrightarrow{W}}{dt}.\overrightarrow{TAS} + gW_z$, avec $\overrightarrow{W}$ le vent et W_z sa composante verticale. De par les équations de Newton, la variation d'énergie spécifique (SER) et l'excès de puissance spécifique (SEP) sont égaux, comme explicité section 2.4.6. Contrairement à l'excès de puissance spécifique (SEP) qui découle du choix d'un modèle des forces ; la variation d'énergie spécifique (SER)

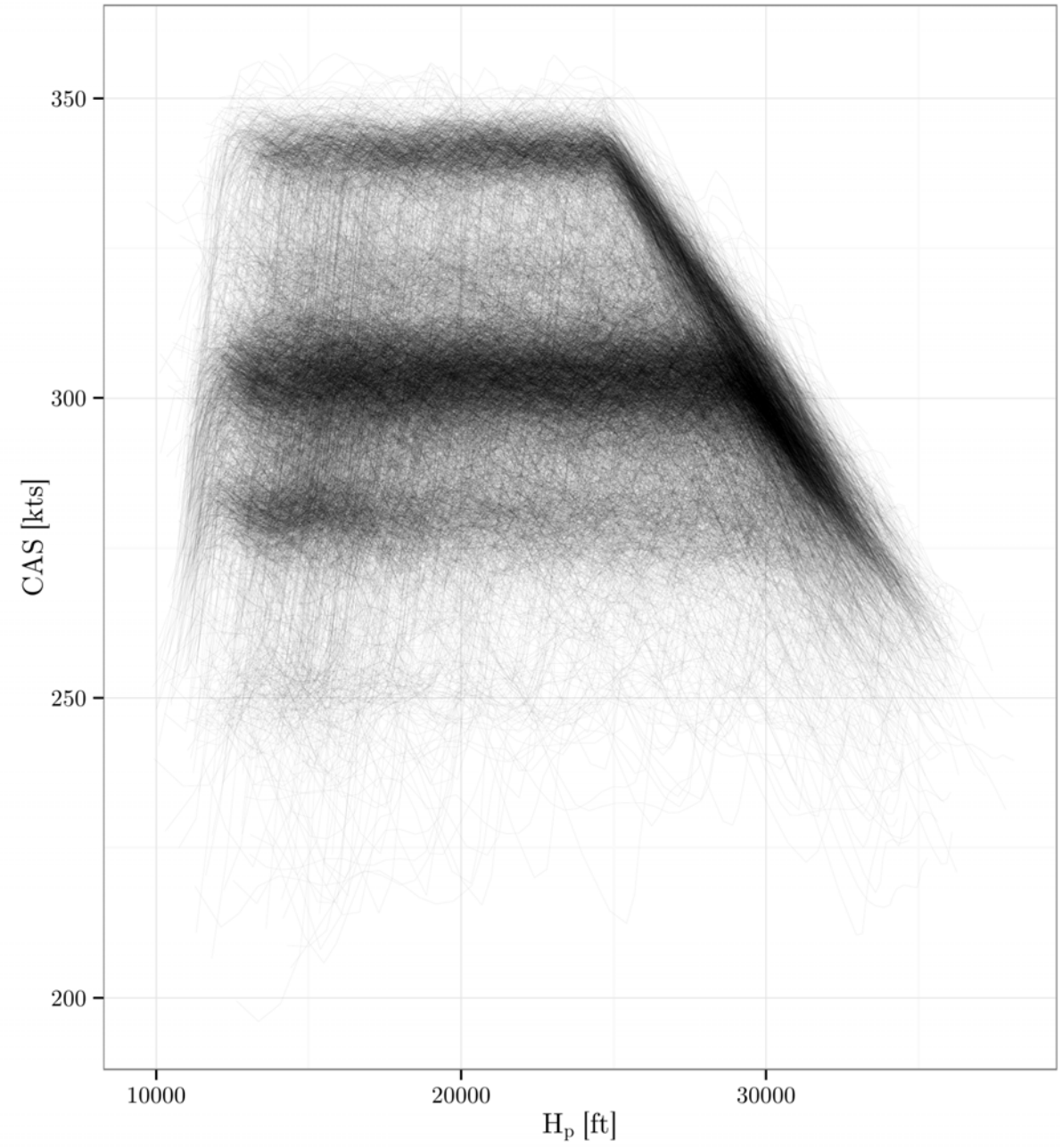


FIGURE 4.12 – Ce graphique représente la vitesse CAS en fonction de l'altitude H_p pour les trajectoires Mode-C.

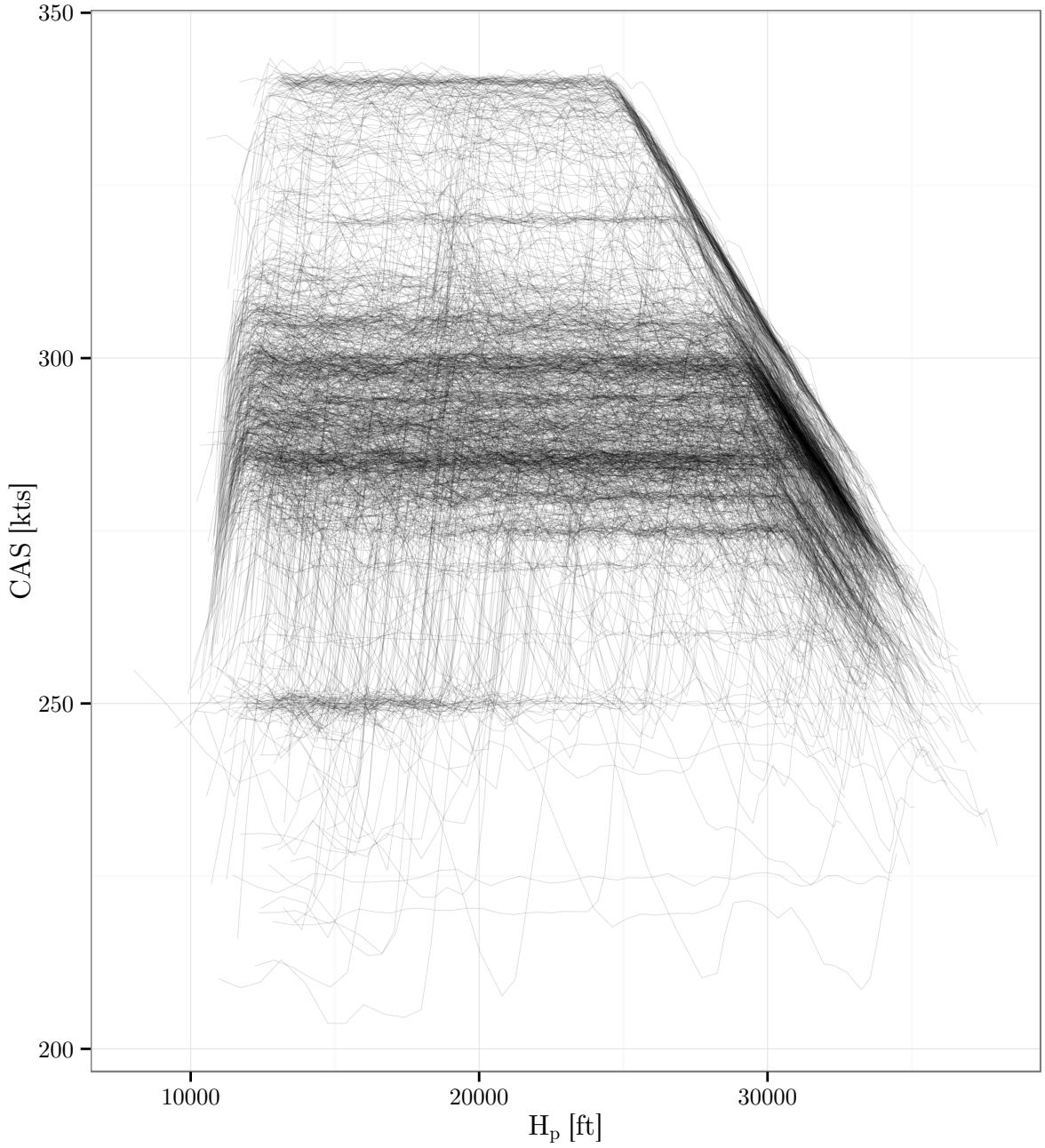


FIGURE 4.13 – Ce graphique représente la vitesse CAS en fonction de l'altitude H_p pour les trajectoires Mode-S.

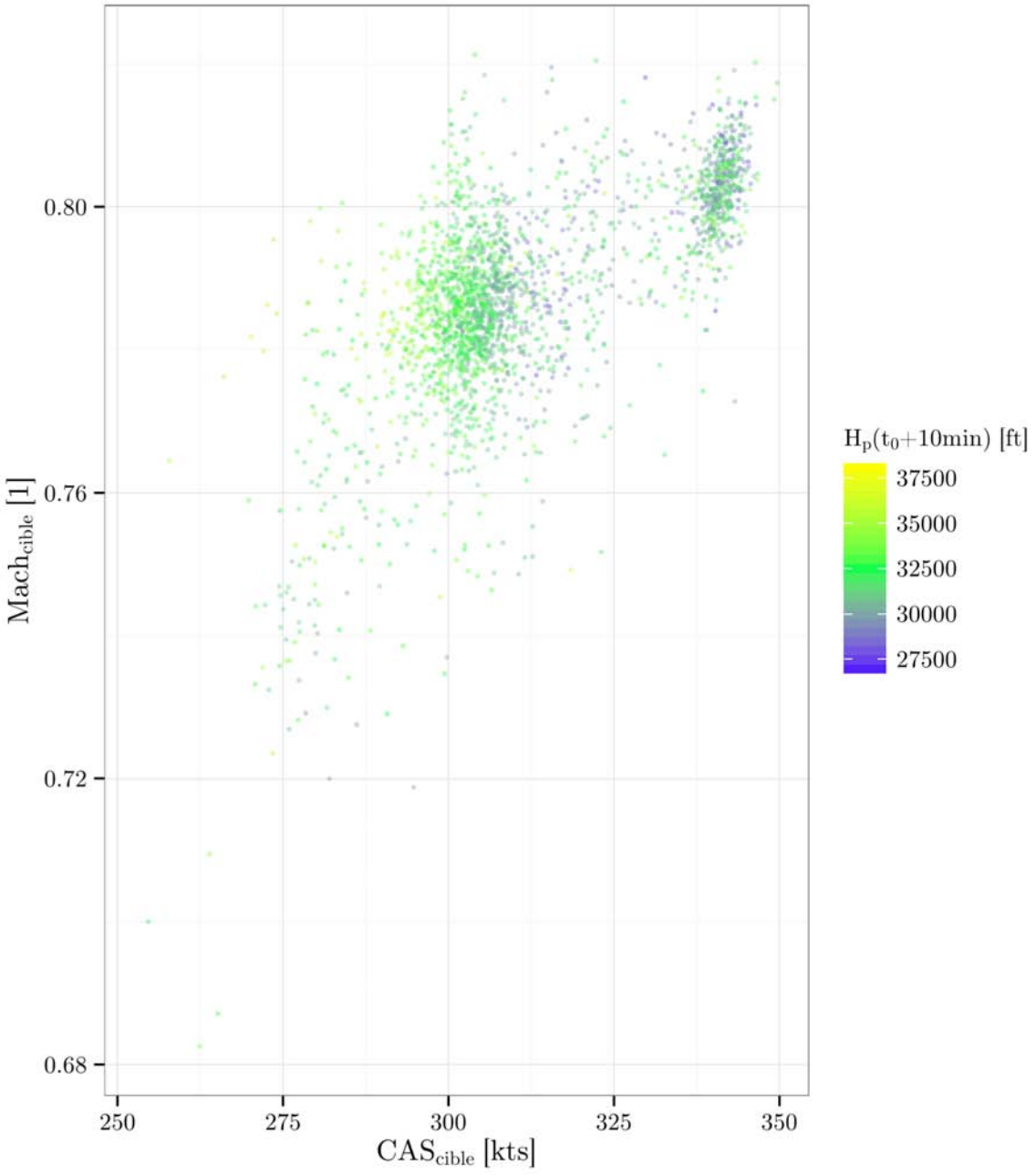


FIGURE 4.14 – Ce graphique représente les différents couples ($CAS_{cible}, Mach_{cible}$) ajusté aux trajectoires Mode-C. Sur cette figure, seule la moitié la mieux ajustée des couples (CAS_{cible},$Mach_{cible}$) est affichée.

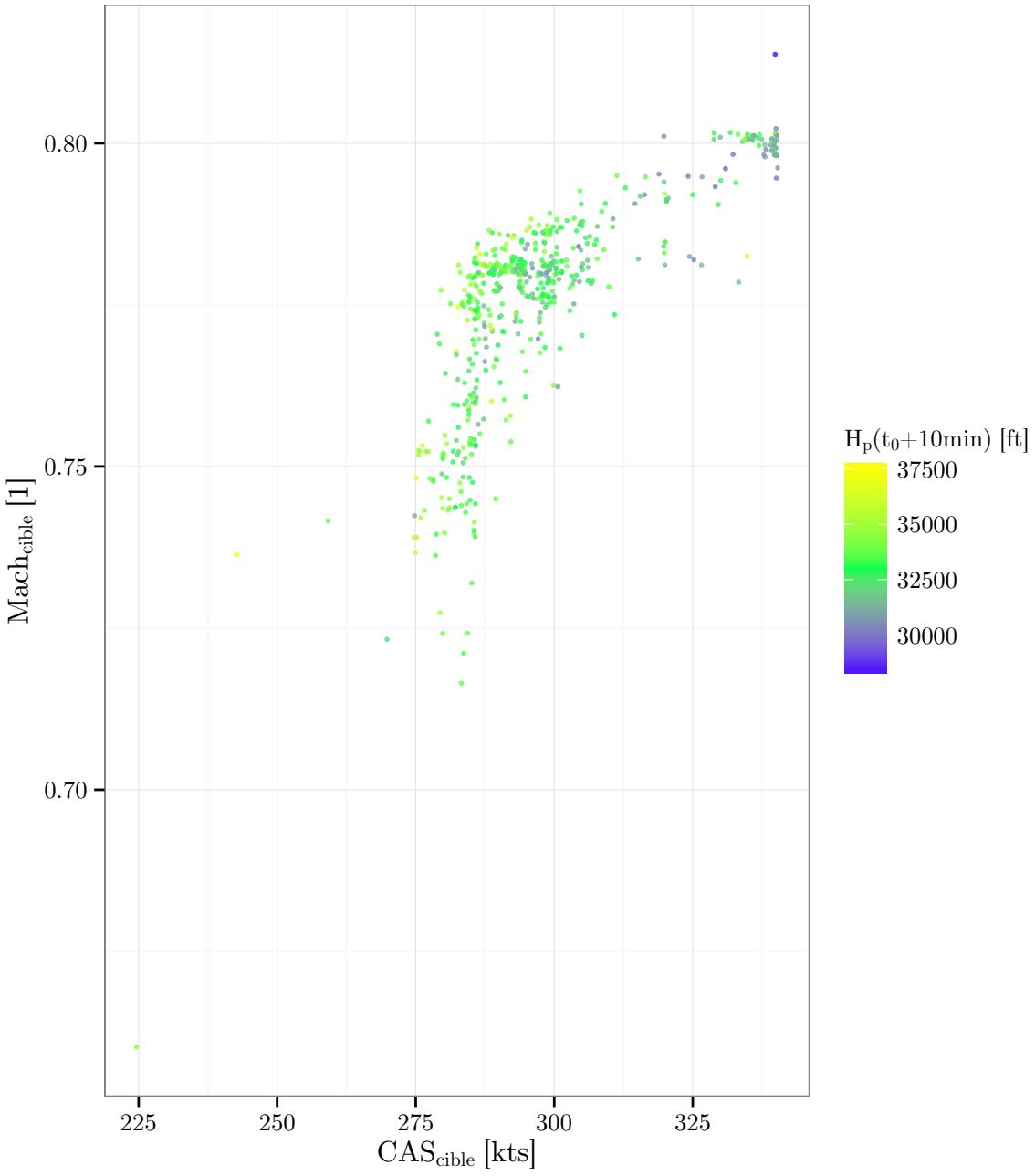


FIGURE 4.15 – Ce graphique représente les différents couples $(CAS_{cible}, Mach_{cible})$ ajusté aux trajectoires Mode-S. Sur cette figure, seule la moitié la mieux ajustée des couples $(CAS_{cible}, Mach_{cible})$ est affichée.

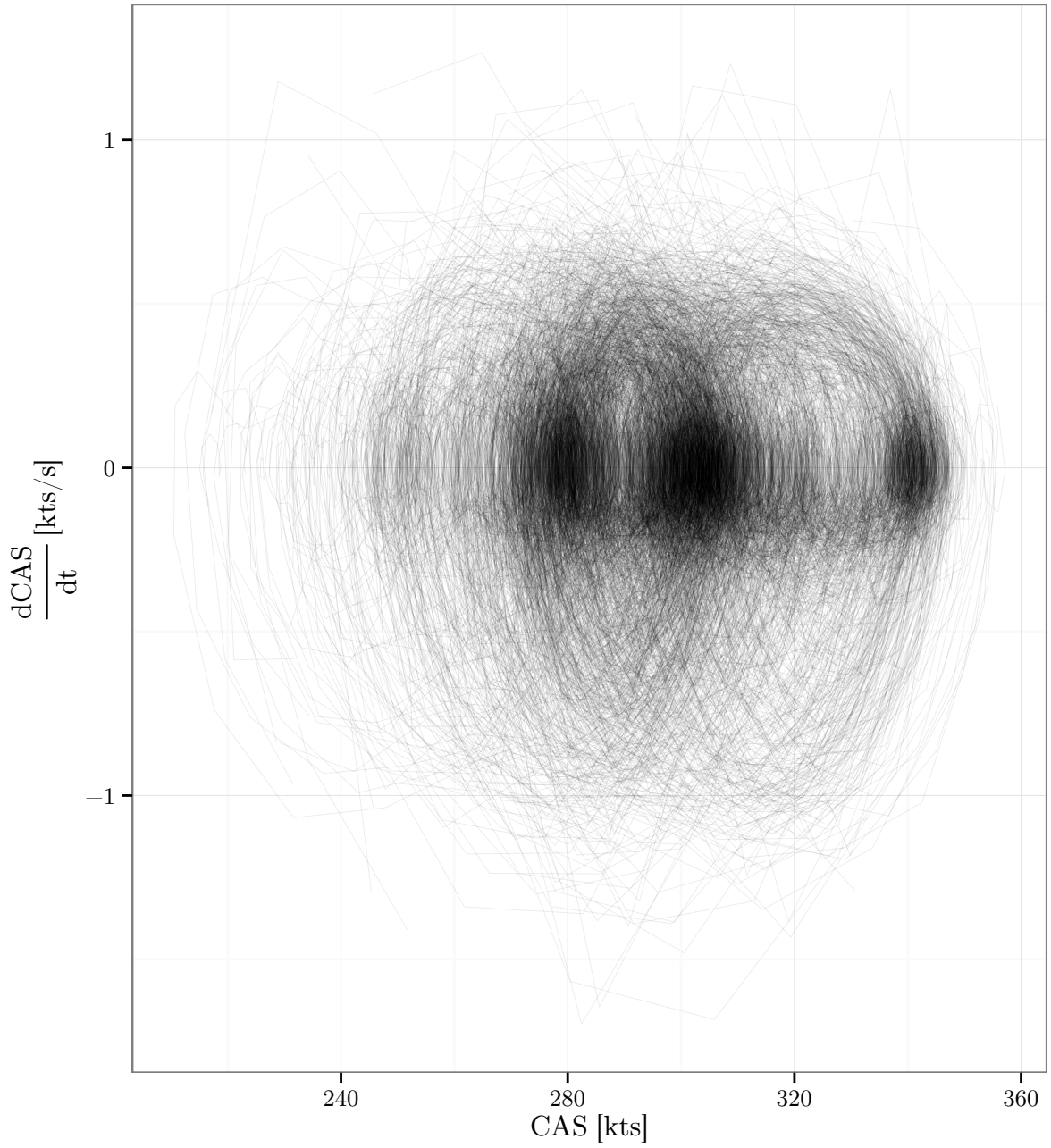


FIGURE 4.16 – Pour les trajectoires Mode-C, ce graphique représente l'accélération en fonction de la vitesse. On a pris les trajectoires dont l'erreur sur la TAS après ajustement d'un profil (CAS_{cible},$Mach_{cible}$) fait partie des 25% les plus élevées et on a uniquement représenté les points d'altitude supérieur à H_{p_0}.

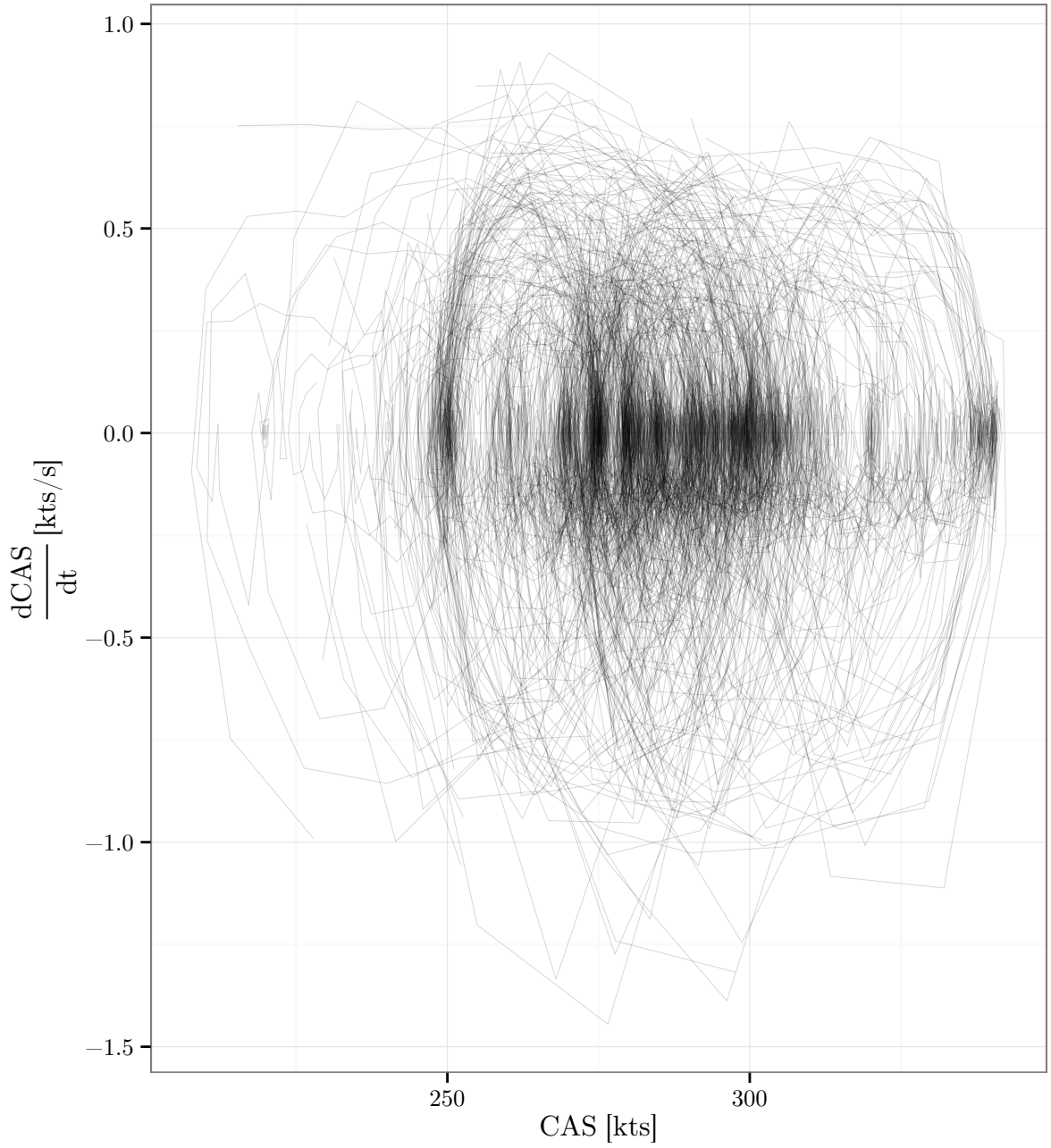


FIGURE 4.17 – Pour les trajectoires Mode-S, ce graphique représente l'accélération en fonction de la vitesse. On a pris les trajectoires dont l'erreur sur la TAS après ajustement d'un profil (CAS_{cible},$Mach_{cible}$) fait partie des 25% les plus élevées et on a uniquement représenté les points d'altitude supérieur à H_{p0}.

peut être calculé uniquement à partir des données disponibles, sans faire intervenir un éventuel modèle des forces.

Ceci va nous permettre de confronter le modèle des forces BADA aux observations. Sur les figures 4.18 et 4.19, on représente la variation d'énergie spécifique (SER) en fonction de l'altitude H_p et de la vitesse CAS. Pour comparer la variation d'énergie spécifique (SER) observée et l'excès de puissance spécifique (SEP) issu du modèle BADA, on a tracé des isolignes d'excès de puissance spécifique correspondant à la modélisation BADA. Le modèle BADA a été utilisé avec une atmosphère en condition ISA sans vent et un avion de type A320 à la masse de référence. Les valeurs de SER et de SEP sont représentées par le biais d'un dégradé de couleurs pour les isolignes et les trajectoires. La correspondance entre les couleurs et les valeurs peut être lues sur les isolignes en W/kg. De l'observation des trajectoires et des valeurs issues du modèle, il ressort que l'altitude H_p a une plus grande importance que la vitesse CAS dans la détermination de la SER. La modélisation faite par BADA de la relation entre le SEP, l'altitude H_p et la vitesse CAS correspond assez bien à la tendance générale émergeant du nuage de trajectoires. Ceci est d'autant plus vrai que la masse utilisée ici dans le modèle BADA est la masse de référence. Les données ne contenant pas la masse, une analyse plus fine sera faite à l'aide de masses estimées dans la section 5. Toutefois, une masse différente n'explique pas les variations de SER que l'on peut observer sur certaines trajectoires.

4.2.4 Profils d'Energy Share Factor

L'ESF est, par définition, égal à $g_0 \frac{dH}{dt} / SER$. Cette grandeur permet de savoir comment est réparti l'excès de puissance spécifique (SEP) entre l'énergie cinétique et l'énergie potentielle. Sur un profil de vitesse de type $(CAS_{cible}, Mach_{cible})$, l'ESF est calculé à partir des formules du manuel utilisateur de BADA ([Nuic 11]). Avec ces formules, le segment CAS constante a un ESF légèrement inférieur à 1 et le segment Mach constant a un ESF supérieur à 1. Ce changement d'ESF a lieu, par définition, à une altitude appelée altitude de conjonction $H_{p,trans}$.

L'ESF calculé sur les trajectoires Mode-C observées va de -170 à 144. Ces grandes valeurs sont une conséquence d'une variation d'énergie spécifique (SER) proche de zéro en certains points.

Dans la section 4.2.2, on a étudié les profils de vitesse de type $(CAS_{cible}, Mach_{cible})$. Parmi ces profils, on a identifié deux valeurs types pour les couples $(CAS_{cible}, Mach_{cible})$. Ces deux valeurs, ($CAS_{cible} = 305$ kts,$Mach_{cible} = 0.78$) et ($CAS_{cible} = 340$ kts,$Mach_{cible} = 0.80$) sont associées respectivement à une altitude de conjonction $H_{p,trans}$ de 28543 ft et de 24660 ft. Dans les figures 4.20 et 4.21, on a tracé l'ESF en fonction de l'altitude H_p. À chaque trajectoire, on a associé une couleur en fonction de leur altitude de conjonction $H_{p,trans}$. On a limité l'amplitude de l'axe vertical à $[0; 3]$. On observe bien le saut d'ESF qui arrive à $H_{p,trans}$. Ceci n'est pas une surprise car la valeur de l'ESF ne dépend que du profil de vitesse. Malgré sa définition faisant intervenir l'excès de puissance spécifique et le taux de montée, l'ESF est uniquement fonction de $\frac{dTAS^2}{dH}$ ([Nuic 11]). Ainsi, il est logique que les valeurs d'ESF soient cohérentes avec l'ajustement des profils de vitesses fait sur les

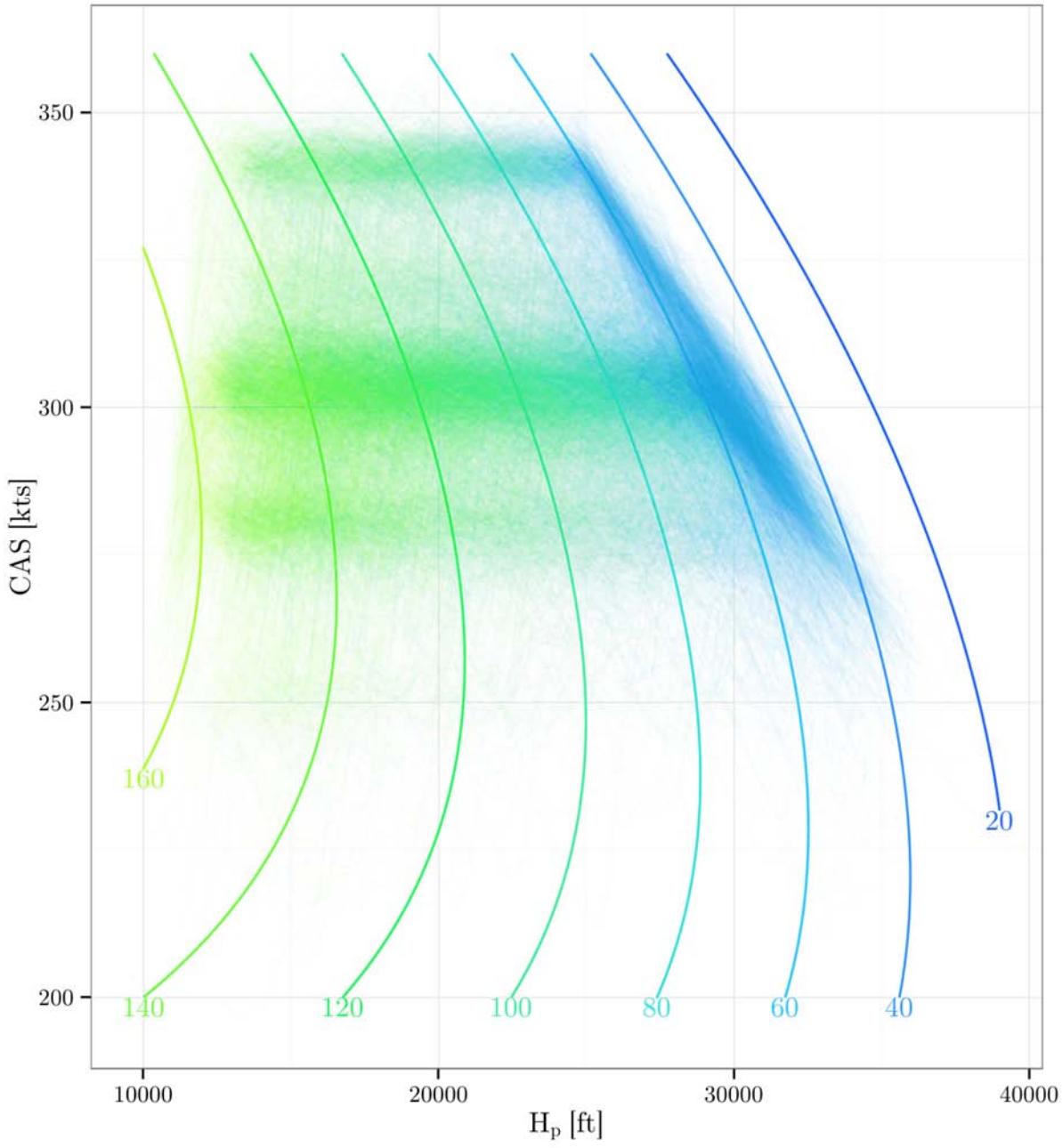


FIGURE 4.18 – On a tracé des isolignes d'excès de puissance spécifique (SEP) correspondant à la modélisation BADA. Pour les trajectoires Mode-C, les valeurs de SER sont représentées par le biais d'un dégradé de couleurs pour les isolignes et les trajectoires. La correspondance entre les couleurs et les valeurs peut être lues sur les isolignes en W/kg.

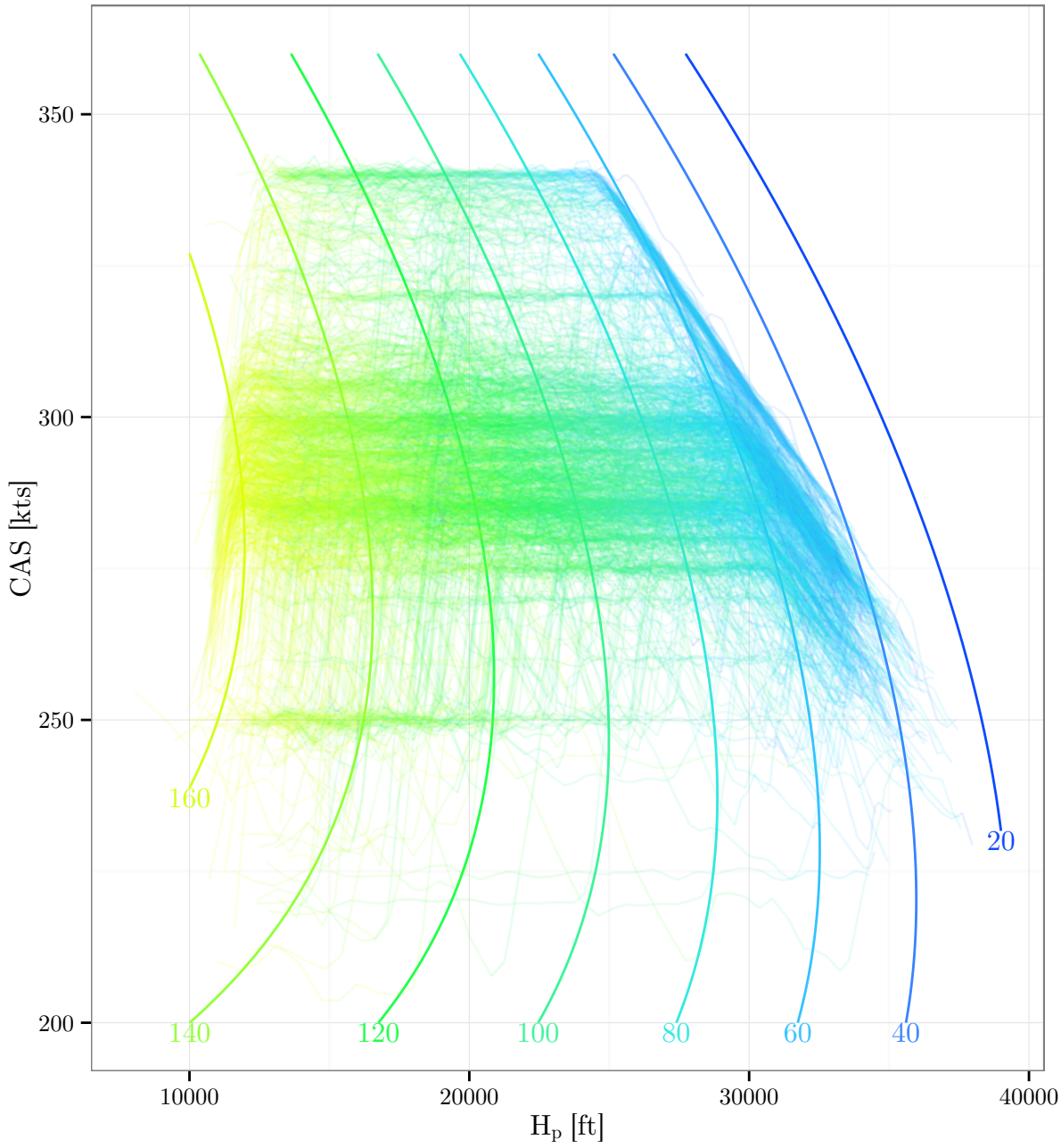


FIGURE 4.19 – Pour les trajectoires Mode-S, on a tracé des isolignes d'excès de puissance spécifique correspondant à la modélisation BADA. Les valeurs de SEP sont représentées par le biais d'un dégradé de couleurs pour les isolignes et les trajectoires. La correspondance entre les couleurs et les valeurs peut être lues sur les isolignes en W/kg.

trajectoires.

4.3 Conclusion

L'analyse des exemples construits permet d'illustrer la difficulté du problème de la prédiction du profil vertical d'un avion en montée.

L'altitude finale dépend de l'excès de puissance spécifique (SEP) appliqué tout au long de la montée. Sa valeur dépend fortement de la masse qui est un paramètre non disponible dans nos données. De plus, pour certaines trajectoires on observe des variations de variation d'énergie spécifique (SER) qui ne sont pas expliquées par le SEP modélisé dans BADA, même en ajustant la masse. Ces variations suggèrent un changement dans la poussée sélectionnée.

L'altitude finale d'un avion est aussi très dépendante du profil de vitesse ; pour nos exemples, une décélération s'accompagne d'un fort taux de montée et inversement, une accélération s'accompagne d'un taux de montée réduit. Ceci est problématique pour les profils exhibant de fortes variations de vitesse. Même pour des profils ayant des variations de vitesse plus faible comme les profils de vitesse de type $(CAS_{cible}, Mach_{cible})$, avoir le couple $(CAS_{cible}, Mach_{cible})$ est important. Ceci est d'autant plus vrai que les valeurs observées de $Mach_{cible}$ couvrent une large plage de valeurs ; or l'altitude finale est très dépendante de $Mach_{cible}$.

Le problème de la prédiction du profil vertical d'un avion en montée est un problème difficile car l'altitude finale dépend fortement des choix faits au cours des 10 minutes de montée à prédire. Au-delà du problème en lui-même, les données utilisées sont issues de mesures lissées. Pour les données Mode-C, les vitesses dans l'air sont issues de l'utilisation des mesures radars et d'une grille météo. Ces données sont ainsi bien moins précises que les données Mode-S qui sont issues de mesures faites par l'équipement à bord de l'avion. Toutefois, le faible nombre d'exemples issues des trajectoires Mode-S peut être problématique pour que des méthodes d'apprentissage artificiel soient appliquées avec succès.

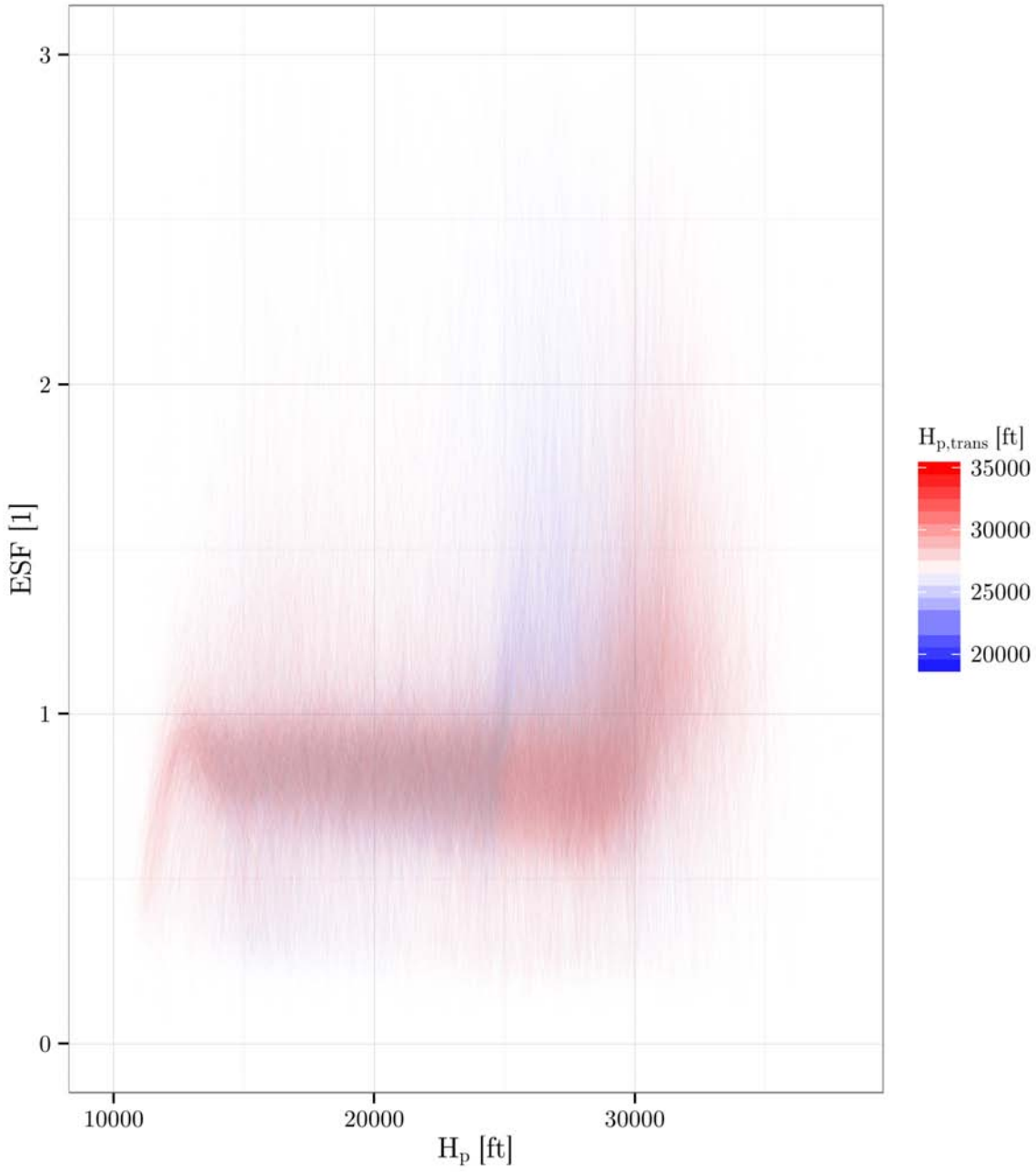


FIGURE 4.20 – Pour les trajectoires Mode-C, ce graphique représente l'ESF en fonction de l'altitude H_p.

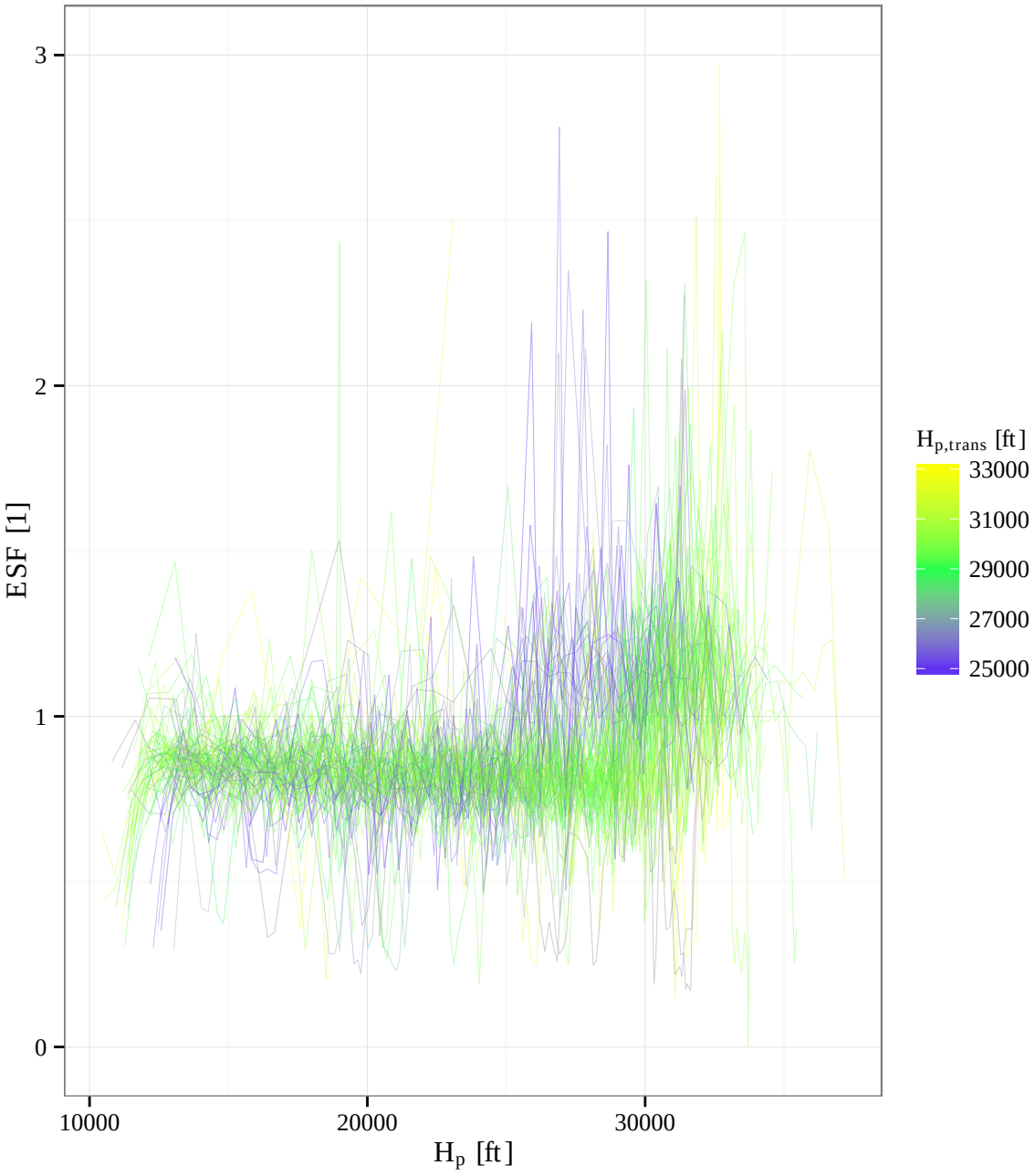


FIGURE 4.21 – Pour les trajectoires Mode-S, ce graphique représente l'ESF en fonction de l'altitude H_p.

Chapitre 5

Estimation de la masse à partir des points passés

Sommaire

5.1 Méthode naïve 92
5.2 Méthode adaptative 95
5.2.1 Ajustement de la masse 95
5.2.2 Adaptation du paramètre de sensibilité 96
5.3 Méthode des moindres carrés 97
5.3.1 Prise en compte de la consommation : relation entre les m_i . . . 98
5.3.2 Approximation de $\mathcal{E}$ 99
5.3.3 Minimisation de $\mathcal{E}_{approx}$ 100
5.4 Méthode robuste 100
5.4.1 Minimisation de $\mathcal{E}_{robuste}$ 101
5.4.2 Newton par intervalles 101
5.5 Comparaison sur des données simulées 101
5.5.1 Ajout du bruit sur les grandeurs observées 103
5.5.2 Résultats 103
5.5.3 Influence de ΔT sur l'erreur 107
5.5.4 Discussion sur les méthodes 109
5.5.5 Limites et conclusion de cette étude 109
5.6 Comparaison sur des données réelles 110
5.6.1 Comparaison de la prédiction 111
5.6.2 Distribution de la masse estimée sur les points passés 112
5.6.3 Discussion sur la masse estimée sur les points futurs 115
5.6.4 Qualité de l'ajustement et qualité de la prédiction 118
5.6.5 Conclusion de l'étude comparative sur les données réelles 121

L'utilisation d'un prédicteur masse-énergie requiert la connaissance de nombreux paramètres. Parmi ceux-ci, la masse est un paramètre important, en particulier en phase de montée. Elle n'est cependant pas disponible dans les systèmes sol actuels.

On peut toutefois essayer d'estimer la masse d'un avion à partir de l'observation de sa trajectoire. Cette idée est exploitée dans différents articles. Dans [Lymperopoulos 06], Lymperopoulos *et al.* modélisent la masse de l'avion et le vent rencontré comme source d'incertitude. Ces variables stochastiques sont échantillonnées suivant une distribution pour générer un jeu de trajectoires aléatoire. Chaque trajectoire est ensuite pondérée selon la probabilité de correspondre à la trajectoire effectivement mesurée. L'incertitude sur les positions futures de l'avion est réduite en sélectionnant les paramètres les plus probables. Cette méthode est uniquement testée par le biais de simulations.

Dans [Slater 02], Slater introduit un mécanisme adaptatif améliorant la prédiction de trajectoire en ajustant dynamiquement la poussée modélisée. L'avion est supposé être à la masse de référence. Les résultats présentés dans [Slater 02] montrent une amélioration significative de la précision sur des trajectoires simulées. Toutefois, cette amélioration n'est plus aussi importante sur les quelques trajectoires réelles étudiées.

Dans [Warren 98], Warren *et al.* proposent d'estimer une masse équivalente. En supposant connu l'*Energy Share Factor*, une masse est ajustée pour que le taux de montée calculé avec cette masse corresponde au taux de montée observé. Les aspects techniques de cet ajustement ne sont toutefois pas abordés. Une seconde étude [Warren 00] suggère d'ajuster la masse en se servant de la variation d'énergie spécifique observée au lieu du taux de montée observé.

Dans des travaux plus récents, [Schultz 12, Thipphavong 13b] introduisent un algorithme qui estime la masse à l'aide d'un mécanisme adaptatif améliorant sa robustesse. Cette méthode a été testée avec succès sur des données simulées et réelles.

Dans ce chapitre, on introduit trois nouvelles méthodes d'estimation de la masse. Ces trois méthodes sont comparées à la méthode adaptative introduite dans [Schultz 12]. Ces méthodes utilisent uniquement le modèle physique pour estimer la masse, comme l'illustre la figure 5.1.

5.1 Méthode naïve

Intuitivement, plus la masse initiale à $t_{initiale}$ est élevée et plus l'altitude à t_{finale} sera basse. Ainsi, on peut envisager d'utiliser plusieurs simulations pour ajuster la masse initiale pour que l'altitude finale prédite par le modèle soit égale à l'altitude finale observée.

Plus formellement, on va rechercher une masse initiale $m_{initiale}$ annulant la fonction $f : m \mapsto \left(H_p^{(pred)}(m) - H_p^{(obs)}\right)(t_{finale})$. En supposant que cette fonction est continue et que $f(m_{max})f(m_{min}) < 0$, on sait que la fonction f s'annule sur $[m_{min}; m_{max}]$. Pour trouver cette masse annulant f, on va pouvoir utiliser une recherche par dichotomie sur l'intervalle $[m_{min}; m_{max}]$. L'utilisation de la recherche dichotomique dans ce cadre est illustrée par la figure 5.2.

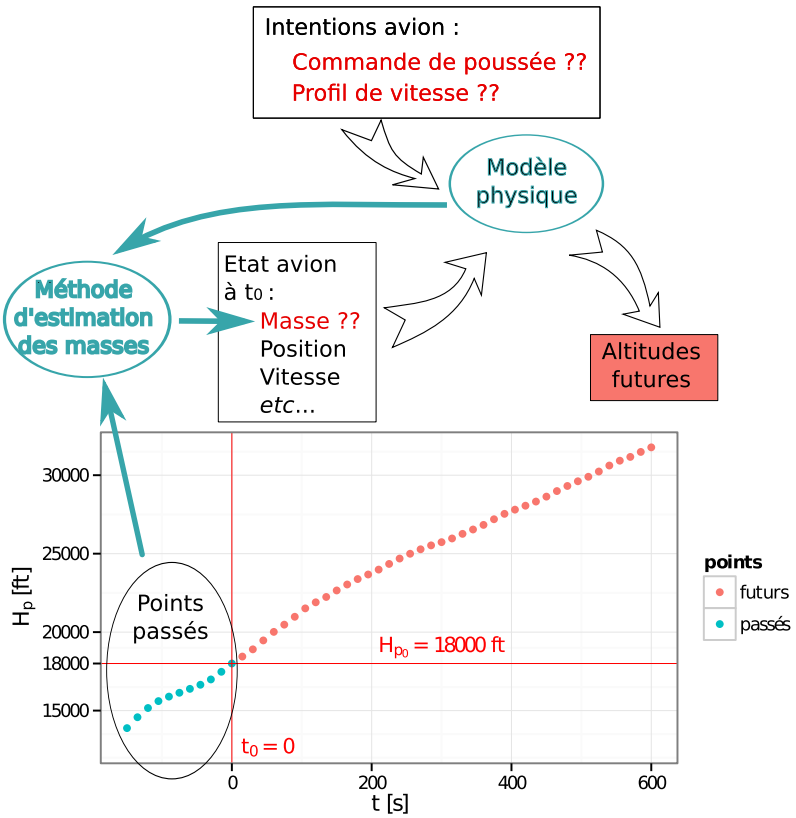


FIGURE 5.1 – Cette figure présente schématiquement l'estimation la masse à partir du modèle physique et de la trajectoire passée.

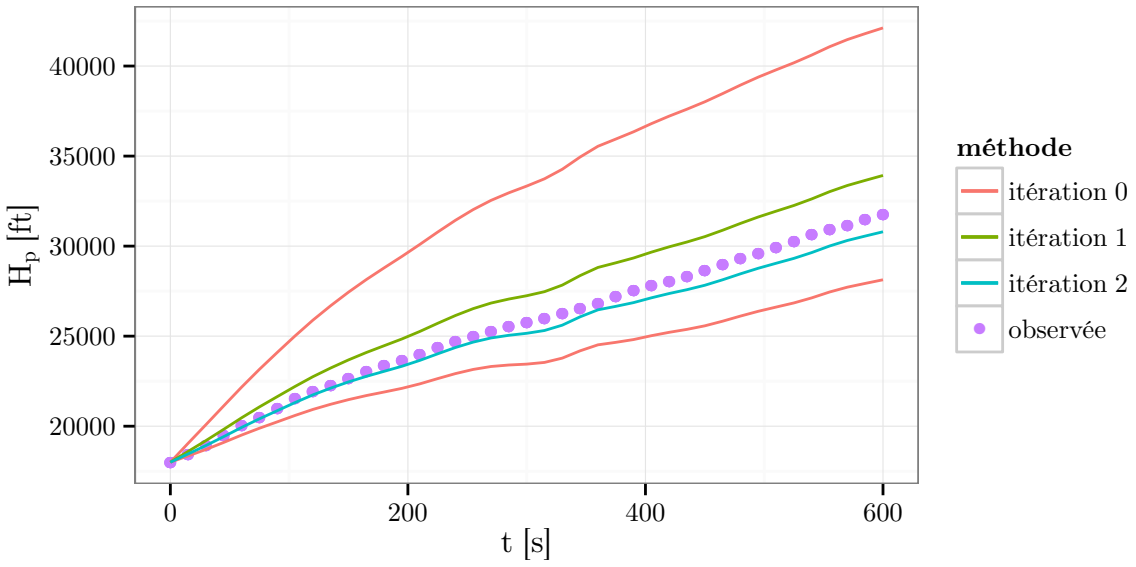


FIGURE 5.2 – Cette figure décrit le déroulement de la méthode naïve pour estimer la masse. À titre d'illustration, la recherche dichotomique est ici appliquée sur les trajectoires simulées de $t_{initiale} = 0$ à $t_{finale} = 600s$ sur cette figure. À l'itération 0, on simule deux trajectoires, une à la masse maximale m_{max} et l'autre à la masse minimale m_{min}. À l'itération 1, on simule une trajectoire avec une masse m_1 qui est la moyenne des masses m_{min} et m_{max}. Comme cette trajectoire est au-dessus de celle observée, la trajectoire simulée à l'itération 2 aura une masse qui sera la moyenne entre m_1 et m_{max}. Cette recherche dichotomique est poursuivie jusqu'à obtenir un encadrement suffisamment étroit de la trajectoire observée.

Cette méthode naïve peut bien sûr être appliquée sur les points passés de la trajectoire avec dans notre cas $t_{initiale} = -150s$ et $t_{finale} = 0$. Cela permet d'utiliser cette masse estimée pour prédire la trajectoire future de $t = 0$ à $t = 600s$. Pour cette méthode, la masse utilisée à $t = 0$ pour prédire le futur est la masse issue de la simulation à $t_{finale} = 0$. Ceci est valable pour toutes les méthodes décrites dans cette section : on estime une masse sur les points passés et on utilise cette masse pour prédire les points futurs.

5.2 Méthode adaptative

La méthode adaptative introduite par Schultz *et al.* dans [Schultz 12] ajuste dynamiquement le poids mg pour que l'excès de puissance spécifique modélisée soit proche de la variation d'énergie spécifique observée. Les points de chaque trajectoire sont considérés séquentiellement selon leur date. La prise en compte d'un point permet de mettre à jour l'estimation du poids. Cette mise à jour dépend d'un paramètre de sensibilité β qui est lui aussi mis à jour à cette occasion en fonction de l'écart entre l'excès de puissance spécifique modélisée et la variation d'énergie spécifique observée. Si cet écart au point courant est bien plus grand que la moyenne des écarts sur les points précédents alors on diminue le paramètre de sensibilité. Dans le cas inverse, on l'augmente. Plus le paramètre de sensibilité est grand, plus l'amplitude de la mise à jour peut être grande. Avec ce mécanisme, des grands écarts isolés vont diminuer le paramètre de sensibilité β et ainsi réduire l'amplitude de la mise à jour de la masse. Ainsi, les valeurs aberrantes par rapport à celles voisines sont détectées et l'adaptation de la masse est faite en conséquence.

On va maintenant décrire plus formellement les deux parties constitutives de cette méthode : l'ajustement du poids et l'adaptation du paramètre de sensibilité. Dans ce document, l'algorithme estime la masse m et non le poids mg. Ceci fait que la présentation de l'algorithme faite dans ce document est légèrement différente de celle faite dans [Schultz 12]. Toutefois, ces deux algorithmes sont équivalents. Pour présenter cet algorithme, on considère une séquence de n points. Dans tout ce qui suit, les observations attachées au point i seront indicées par i.

5.2.1 Ajustement de la masse

Au point i, l'avion a une masse qui vaut m_i. Celle-ci est estimée par la masse $\hat{m}_i$. Pour la concision des formules, on note $Q_i = \left(g_0 \frac{T}{T_{ISA}} \frac{dH_p}{dt} + V_a \frac{dV_a}{dt} - gW_Z + \dot{\overrightarrow{W}}.\overrightarrow{V_a}\right)_i$, $Puiss_i(m) = (Thr_i - D_i(m)) V_{ai}$ et $P_i(m) = Q_i m - Puiss_i(m)$. L'égalité entre l'excès de puissance

spécifique P_{es} et la variation d'énergie spécifique E_s s'écrit :

$$\begin{aligned}
&P_{esi} = E_{si} \\
\Leftrightarrow &\frac{Thr_i - D_i\left(m_i\right)}{m_i} V_{ai} + \left(gW_Z - \dot{\overrightarrow{W}}.\overrightarrow{V_a}\right)_i = \left(g_0 \frac{T}{T_{ISA}} \frac{dH_p}{dt} + V_a \frac{dV_a}{dt}\right)_i \\
\Leftrightarrow &\frac{Puiss_i\left(m_i\right)}{m_i} = \left(g_0 \frac{T}{T_{ISA}} \frac{dH_p}{dt} + V_a \frac{dV_a}{dt} - gW_Z + \dot{\overrightarrow{W}}.\overrightarrow{V_a}\right)_i \\
\Leftrightarrow &\frac{Puiss_i\left(m_i\right)}{m_i} = Q_i \\
\Leftrightarrow &\frac{P_i\left(m_i\right)}{m_i} = 0
\end{aligned} \tag{5.1}$$

Ainsi, l'équation (5.1) est équivalente à $Q_i = \frac{Puiss_i(m_i)}{m_i}$.

Dans cette méthode, les forces au point i sont calculées en utilisant la masse $\hat{m}_{i-1}$ estimée lors de la prise en compte du $(i-1)^e$ point. En approximant $Puiss_i\left(m_i\right)$ par $Puiss_i\left(\hat{m}_{i-1}\right)$ et m_i par $\hat{m}_i$, on obtient l'équation suivante :

$$\begin{aligned}
\hat{m}_i =& \frac{Puiss_i(\hat{m}_{i-1})}{Q_i} \\
=& \frac{Puiss_i(\hat{m}_{i-1})}{\frac{Puiss_i(\hat{m}_{i-1})}{\hat{m}_{i-1}} + (Q_i - \frac{Puiss_i(\hat{m}_{i-1})}{\hat{m}_{i-1}})} \\
=& \frac{1}{\frac{1}{\hat{m}_{i-1}} + \frac{1}{Puiss_i(\hat{m}_{i-1})} \underbrace{\left(Q_i - \frac{Puiss_i(\hat{m}_{i-1})}{\hat{m}_{i-1}}\right)}_{\text{erreur faite en utilisant } \hat{m}_{i-1} \text{ pour calculer } E_{si}}} \\
=& \hat{m}_{i-1}\left(1 + \frac{P_i(\hat{m}_{i-1})}{Puiss_i(\hat{m}_{i-1})}\right)^{-1}
\end{aligned} \tag{5.2}$$

Pour les raisons expliquées précédemment, [Schultz 12] introduisent un paramètre de sensibilité β dans le coefficient de mise à jour de l'équation (5.2). Finalement, la masse est mise à jour par l'équation :

$$\hat{m}_i = \hat{m}_{i-1}\left[1 + \beta_i \frac{P_i(\hat{m}_{i-1})}{Puiss_i(\hat{m}_{i-1})}\right]^{-1} \tag{5.3}$$

5.2.2 Adaptation du paramètre de sensibilité

Le paramètre de sensibilité β_i est adapté en fonction de la comparaison entre les écarts sur le point courant et les points précédents. En posant $\Delta\dot{E}_i = \frac{P_i(\hat{m}_{i-1})}{\hat{m}_{i-1} g V_a}$, la règle d'adapta-

tion donnée par [Schultz 12] est la suivante :

$$\begin{aligned} &\text{si } i > 0 \text{ et } \Delta\dot{E}_i > 0.0001 \text{ et } \left|\frac{\Delta\dot{E}_i - \Delta\dot{E}_{avg}}{\Delta\dot{E}_{avg}}\right| < 3 \\ &\quad \text{alors} \\ &\qquad \beta_i = max(0.205, \beta_{i-1} + 0.05) \\ &\quad \text{sinon} \\ &\qquad \beta_i = 0.005 \end{aligned} \tag{5.4}$$

Dans l'équation (5.4), $\Delta\dot{E}_{avg}$ est la moyenne des $\Delta\dot{E}_i$ sur les cinq points précédents. Sur les premiers points de la trajectoire, cette moyenne est calculée sur tous les points disponibles lorsque leur nombre ne dépasse pas cinq.

Avec ce mécanisme, si les $\Delta\dot{E}_i$ sont grands, de valeurs similaires, alors β_i augmente et renforce ainsi l'adaptation. Dans le cas contraire, β_i est petit. Ainsi, les valeurs isolées de $\Delta\dot{E}_i$ auront un impact limité sur la masse estimée. Ceci améliore la robustesse du processus.

L'algorithme commence avec une masse initiale $\hat{m}_0$ (typiquement la masse de référence donnée par le modèle BADA). À chaque itération, la variation de la masse estimée est bornée : dans nos expérimentations, elle est limitée à 2,5 %[1] de la masse de référence. Au cours de tout le processus d'estimation, la masse estimée est bornée ; elle est maintenue entre 80 % et 120 % de la masse de référence.

5.3 Méthode des moindres carrés

Dans la méthode adaptative évoquée précédemment, la masse estimée est mise à jour séquentiellement à chaque nouveau point considéré. L'algorithme commence avec une estimation *a priori* $\hat{m}_0$ et termine avec une estimation $\hat{m}_n$ après n itérations, n étant le nombre de points considérés. Cette méthode utilise les équations (5.1) pour obtenir une relation de récurrence sur les $\hat{m}_i$.

On peut exploiter les équations (5.1) différemment. Par exemple, on peut résoudre pour chaque point i l'équation $P_i(m_i) = 0$. On obtient ainsi une séquence de masses qui peut ne pas être réaliste, comme illustré par la figure 5.3. En effet, dans la réalité, les masses m_i ne sont pas indépendantes les unes des autres. Par exemple, comme elles sont ordonnées selon leur date, la séquence de masses doit être décroissante. Compte tenu des incertitudes et des erreurs d'observations, les équations (5.1) ne vont pas être vérifiées en tout point i.

Dans ce cadre, nous ([Alligier 12]) reprenons l'idée évoquée dans [Warren 00] : la masse estimée est celle minimisant la somme des écarts au carré entre la variation d'énergie spécifique et l'excès de puissance spécifique modélisé. Dans la méthode des moindres carrés que nous proposons, les points sont considérés tous en même temps. L'erreur totale à

1. Cette valeur est différente de celle donnée dans [Schultz 12], mais elle donne de meilleurs résultats sur nos données.

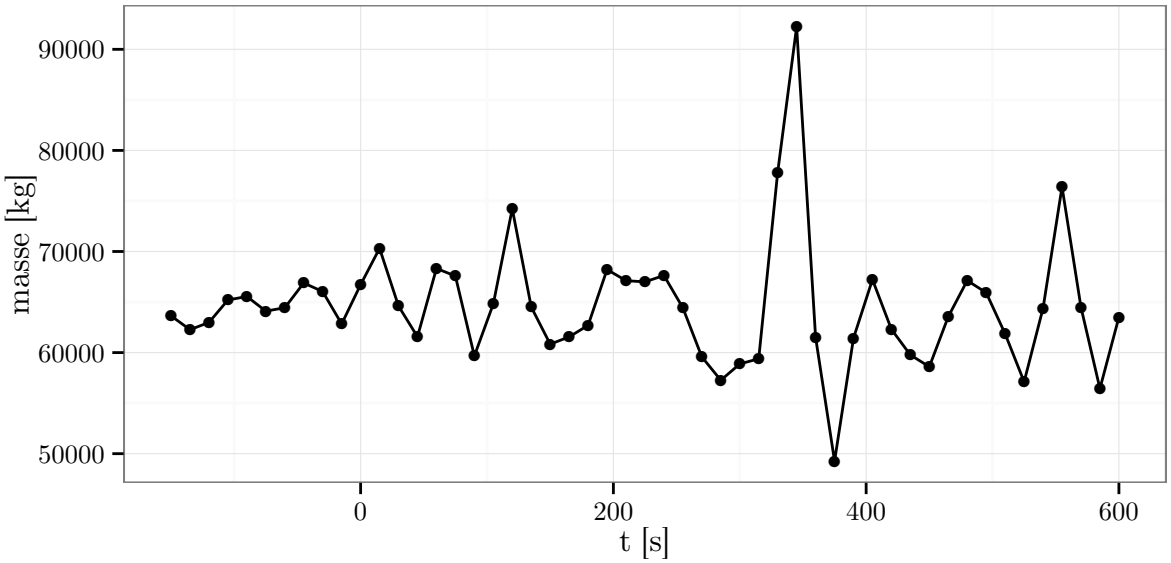


FIGURE 5.3 – Cette figure présente, pour une trajectoire, la masse estimée à chaque point i obtenue par résolution de chaque équation $P_i(m_i) = 0$.

minimiser $\mathcal{E}$ sur les n points est la suivante :

$$\mathcal{E}(m_1, \ldots, m_n) = \sum_{i=1}^{n} \left(\frac{Puiss_i(m_i)}{m_i} - Q_i \right)^2 = \sum_{i=1}^{n} \left(\frac{P_i(m_i)}{m_i} \right)^2 \tag{5.5}$$

Dans l'équation (5.5), l'erreur à minimiser est liée à l'excès de puissance spécifique P_{es} modélisé et non la puissance modélisée. Ce choix est motivé par l'utilisation de la masse estimée ; elle sert à calculer la trajectoire prédite qui est issue de l'intégration de l'équation (2.28). Seule la trajectoire, i.e. la séquence de couples (H_p, V_a), nous intéresse. Or, seul le ratio de la puissance sur la masse est important pour cela : une puissance double avec une masse double générera une trajectoire identique à celle générée par une puissance et une masse simple. Cette question est également évoquée dans la section 2.6.5.

5.3.1 Prise en compte de la consommation : relation entre les m_i

La variation de la masse de l'avion est causée par la consommation de carburant. Dans le modèle BADA, cette consommation est modélisée par l'équation (2.30).

Avec cette équation, la connaissance de la température, de l'altitude et de la vitesse

permet d'écrire la masse m_i en fonction de m_n.

$$m_i = m_n + \int_{t_i}^{t_n} f_{nom}(V_a(t), H_p(t), \Delta T(t))dt \tag{5.6a}$$

$$\simeq m_n + \sum_{k=i}^{n-1} \frac{f_{nom}(t_{k+1}) + f_{nom}(t_k)}{2}(t_{k+1} - t_k) \tag{5.6b}$$

$$= m_n + \delta_i \tag{5.6c}$$

La quantité $\delta_i = \sum_{k=i}^{n-1} \frac{f_{nom}(t_{k+1})+f_{nom}(t_k)}{2}(t_{k+1} - t_k)$ peut être calculée à partir des données qui sont disponibles en chaque point i. Ainsi, la somme des erreurs au carré $\mathcal{E}$ peut être réécrite :

$$\mathcal{E}(m_n) = \sum_{i=1}^{n} \left(\frac{\tilde{P}_i(m_n)}{(m_n + \delta_i)} \right)^2 \tag{5.7a}$$

$$\text{avec } \tilde{P}_i(m_n) = P_i(m_n + \delta_i) \tag{5.7b}$$

Au final, en prenant en compte la consommation, l'équation (5.6c) permet d'obtenir l'équation (5.7a) pour laquelle $\mathcal{E}$ est uniquement fonction de m_n.

5.3.2 Approximation de $\mathcal{E}$

La masse estimée de l'avion est la masse $\hat{m}_n$ minimisant $\mathcal{E}$. Avec le modèle BADA, P_i et $\tilde{P}_i$ sont des polynômes du second degré. La somme des erreurs $\mathcal{E}$ est ainsi une fonction rationnelle (i.e. un ratio de fonctions polynomiales). Le minimum $\hat{m}_n$ de cette fonction sur $]0; +\infty[$ satisfait l'équation $\mathcal{E}'(m) = 0$. Quand on réduit au même dénominateur cette équation, résoudre $\mathcal{E}'(m) = 0$ se ramène à trouver les racines d'un polynôme de degré au plus $3(n-1)+4$. Trouver les racines d'un polynôme de degré élevé est une tâche difficile à cause de problèmes liés aux approximations numériques ([Wilkinson 94]). Pour contourner ces problèmes, on approxime $\mathcal{E}$ par $\mathcal{E}_{approx}$ définie par les équations ci-dessous :

$$\delta_{avg} = \frac{1}{n} \sum_{i=1}^{n} \delta_i \tag{5.8a}$$

$$m_{avg}(m_n) = m_n + \delta_{avg} \tag{5.8b}$$

$$\mathcal{E}_{approx}(m_n) = \sum_{i=1}^{n} \left(\frac{\tilde{P}_i(m_n)}{m_{avg}(m_n)} \right)^2 \tag{5.8c}$$

Par cette définition, chaque terme $\frac{\tilde{P}_i(m_n)}{m_n+\delta_i}$ est approximé par $\frac{\tilde{P}_i(m_n)}{m_n+\delta_{avg}}$. L'écart relatif entre ces deux quantités est facilement calculable :

$$\frac{\tilde{P}_i(m_n)}{m_n + \delta_{avg}} = \left[1 + \frac{\delta_i - \delta_{avg}}{m_n + \delta_{avg}} \right] \frac{\tilde{P}_i(m_n)}{m_n + \delta_i} \tag{5.9}$$

Comme la variation relative de la masse est faible dans la fenêtre temporelle considérée, l'erreur relative due à l'approximation est également faible. Nous montrons dans [Alligier 13a] que l'utilisation de $\mathcal{E}_{approx}$ sur une trajectoire simulée, non « bruitée », permet de retrouver, à la centaine de grammes près, la masse avec laquelle la trajectoire a été générée.

5.3.3 Minimisation de $\mathcal{E}_{approx}$

Avec cette approximation, la dérivée de la fonction d'erreur est donnée par l'équation ci-dessous :

$$\mathcal{E}'_{approx}(m_n) = \frac{2}{(m_{avg}(m_n))^3} \sum_{i=1}^{n} \tilde{P}_i(m_n) \left[\tilde{P}_i{}'(m_n) m_{avg}(m_n) - \tilde{P}_i(m_n) m_{avg}{}'(m_n) \right] \quad (5.10)$$

Ainsi, la masse estimée $\hat{m}_n$ est racine du polynôme de degré 4 défini par l'équation (5.11) ci-dessous :

$$S(\hat{m}_n) = \sum_{i=1}^{n} \tilde{P}_i(\hat{m}_n) \left[\tilde{P}_i{}'(\hat{m}_n) m_{avg}(\hat{m}_n) - \tilde{P}_i(\hat{m}_n) m_{avg}{}'(\hat{m}_n) \right] = 0 \quad (5.11)$$

Cette équation peut être résolue analytiquement en utilisant la méthode de Ferrari. Toutefois, une résolution analytique peut donner des racines éloignées des vraies racines du fait des approximations numériques inhérentes aux calculs sur ordinateur [Higham 02]. Pour trouver les racines d'un polynôme, on a utilisé une méthode numérique utilisant la décomposition QR de la matrice compagnon du polynôme considéré [Edelman 95, Martin 71]. Cette méthode est implémentée dans la *GNU Scientific Library*. Dans nos expérimentations, cette méthode s'est révélée aussi rapide que la méthode de Ferrari. Parmi les quatre solutions réelles potentiellement isolées, on sélectionne celle minimisant $\mathcal{E}_{approx}$. Cette dernière sera notre masse estimée $\hat{m}_n$. À partir de celle-ci et des consommations δ_i, on peut reconstituer une séquence de masse estimée cohérente avec la modélisation de la consommation de carburant faite dans BADA.

5.4 Méthode robuste

Cette méthode est une variante de la méthode des moindres carrés. En effet, on travaille toujours avec une somme d'erreurs auxquelles on associe une perte. Ici, on remplace la perte quadratique de la méthode précédente par une perte pénalisant moins les grandes erreurs. Avec cette méthode, on remplace la fonction de perte quadratique par une fonction de perte pour laquelle la pénalité augmente moins vite lorsqu'on s'éloigne de zéro. Ceci permet de donner moins de poids aux valeurs aberrantes.

$$\mathcal{E}_{robuste}(m_n) = \sum_{i=1}^{n} L_r \left(\frac{\tilde{P}_i(m_n)}{m_{avg}(m_n)} \right) \quad (5.12a)$$

$$\text{avec } L_r(\varepsilon) = \left(\sqrt{1 + \frac{\varepsilon^2}{r}} - 1 \right) r \quad (5.12b)$$

5.4.1 Minimisation de $\mathcal{E}_{robuste}$

La fonction de perte pseudo-Huber L_r ([Hartley 03], page 619) est C^∞. Ainsi, la masse m minimisant $\mathcal{E}_{robuste}$ sur $]-\delta_{avg}; +\infty[$ satisfait l'équation $\mathcal{E}'_{robuste}(m) = 0$. On va donc rechercher la masse minimisant $\mathcal{E}_{robuste}$ parmi les solutions de l'équation $\mathcal{E}'_{robuste}(m) = 0$. Pour résoudre cette équation, on s'appuie sur l'algorithme de Newton par intervalles ([Hansen 83]). Cet algorithme, popularisé par [Van Hentenryck 97], est une extension de la méthode de Newton à l'analyse par intervalle introduite par [Moore 66]. Pour tout ce qui suit, la méthode robuste est employée avec $r = 30$ W/kg.

5.4.2 Newton par intervalles

L'algorithme de Newton par intervalles permet de trouver et borner rigoureusement tous les zéros d'une fonction dérivable f sur un intervalle de départ X. Notons F (respectivement F') une fonction d'inclusion de f (respectivement f'). Pour tout $c \in X$, on définit l'opérateur de Newton par :

$$N(X, c) = c - \frac{F(c)}{F'(X)}. \tag{5.13}$$

On construit par récurrence la suite d'intervalles $X_0 = X$ et $X_{k+1} = X_k \cap N(X_k, c_k), \forall k \geq 0$. Alors tous les zéros de f sont automatiquement séparés et bornés en utilisant l'arithmétique par intervalles. Cette séparation des zéros intervient quand $X_k \cap N(X_k, c_k)$ est l'union de deux intervalles. Chacun de ces deux intervalles est susceptible de contenir un zéro ; on applique l'algorithme de Newton par intervalles à chacun de ces deux intervalles. De plus, $X_{k+1} = \{\}$ implique qu'il n'existe aucun zéro de f dans X_k, et $X_{k+1} \subset \mathring{X}_k$ prouve l'existence et l'unicité d'un zéro dans X_{k+1}. Si $0 \notin F'(X_k)$, ce qui implique f monotone, alors on a une convergence q-quadratique de l'algorithme de Newton par intervalles. Si on a $X_{k+1} = X_k$ alors on a atteint un point fixe et l'espace X_k est scindé en deux parties auxquelles on applique l'algorithme de Newton par intervalles. Il est à noter que l'arithmétique par intervalles prend en compte les erreurs d'arrondis dues à l'utilisation de l'arithmétique flottante. Une itération de cet algorithme est illustrée par la figure 5.4.

5.5 Comparaison sur des données simulées

Les méthodes d'estimations de masses présentées précédemment ont été comparées sur des jeux de trajectoires simulées. Avec ces jeux de trajectoires, on pourra évaluer la robustesse des différentes méthodes d'estimation de la masse en ajoutant du bruit aux variables observées. Ces trajectoires d'avions en phase de montée sont générées avec la version 3.9 du modèle BADA. On a considéré trois types avions différents : l'A320 qui est un court-courrier, l'A333 qui est un moyen-courrier et le B744 qui est un long-courrier. La méthode naïve est écartée de l'étude car les calculs qu'elle engendre sont trop longs.

On a choisi de simuler des trajectoires à poussée *max climb*. Les profils de vitesse utilisés sont de type $(CAS, Mach)$ i.e. l'avion vole à une vitesse CAS égale à CAS jusqu'à l'altitude

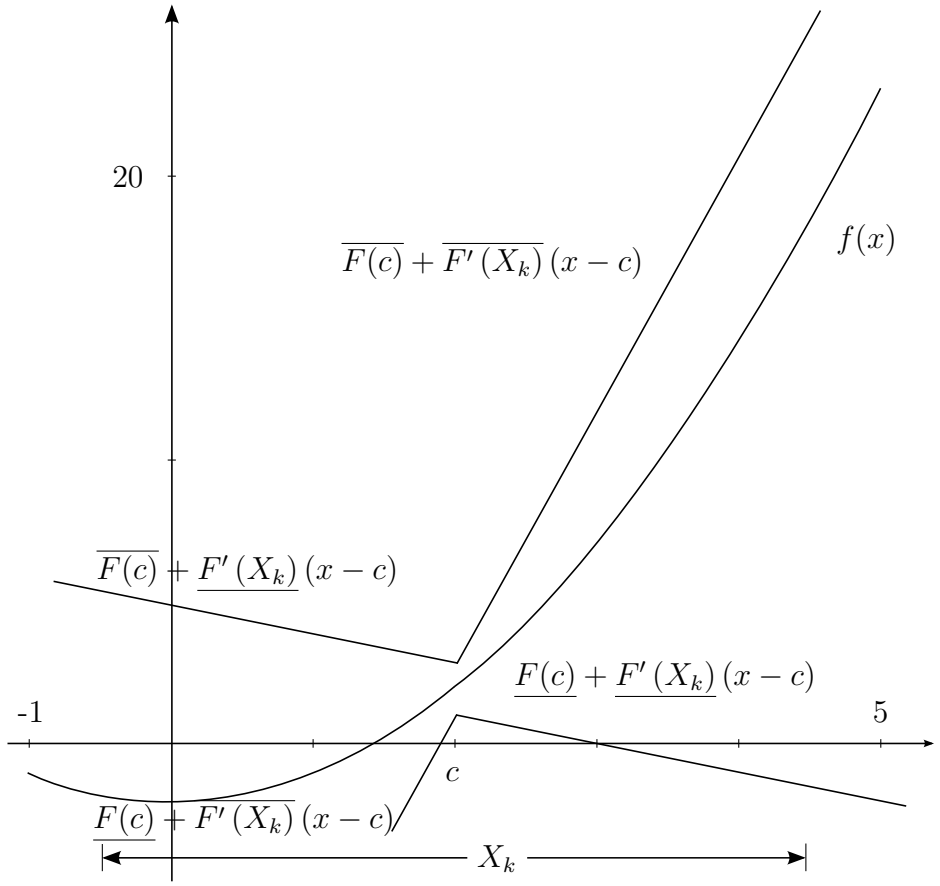


FIGURE 5.4 – Cette figure illustre une itération de l'algorithme de Newton par intervalle avec $f(x) = x^2 - 2$ et $X_k = [-\frac{1}{2}, \frac{9}{2}]$. Dans cet exemple, on suppose que $F(c)$ est un intervalle relativement grand ce qui permet de mieux visualiser les deux cônes obtenus. Graphiquement, les zéros sont sur l'intersection de ces cônes et de l'axe des x. Ainsi, on a $X_{k+1} = [-\frac{1}{2}; \frac{17}{9}] \cup [3; \frac{9}{2}]$. Il faut répéter cette opération sur ces deux intervalles.

de conjonction puis vole à une vitesse Mach égale à $Mach$ au dessus de cette altitude. Les atmosphères utilisées sont celles décrites dans la section 2.2 ; elles sont paramétrées par ΔT. Différentes valeurs de ΔT sont utilisées pour simuler une atmosphère chaude ou froide. Au final, les trajectoires générées sont caractérisées par le quadruplet $(CAS, Mach, \Delta T, mass)$.

Les jeux de trajectoires sont générés en tirant, indépendamment les uns des autres, chacun des quatre paramètres $(CAS, Mach, \Delta T, mass)$ selon une loi uniforme comme décrit dans le tableau 5.1. Une telle distribution n'est pas réaliste mais elle permet d'évaluer nos méthodes d'estimation de la masse sur une large variété de trajectoires. Chaque trajectoire contient 11 points espacés de 15 s. En chacun de ces points, on observe les variables T, H_p, V_a, $\frac{dH_p}{dt}$ et $\frac{dV_a}{dt}$. Par construction, le onzième point de chaque trajectoire se situe à 18 000 ft. Les jeux de trajectoires considérés contiennent 1 000 trajectoires chacun.

TABLE 5.1 – Ce tableau présente la distribution des paramètres utilisés pour générer les trajectoires.

paramètres	distribution
CAS	$CAS_{ref} + uniforme([-30; 30])$
Mach	$Mach_{ref} + uniforme([-0.03; 0.03])$
ΔT	$uniforme([-20; 20])$
mass	$mass_{ref} \times uniforme([0.8; 1.2])$

5.5.1 Ajout du bruit sur les grandeurs observées

Sur les trajectoires observées dans la réalité, les grandeurs T, H_p, V_a, $\frac{dH_p}{dt}$ et $\frac{dV_a}{dt}$ sont entachées d'erreurs. Avec nos trajectoires simulées, on ajoute un bruit sur ces cinq grandeurs ce qui permet d'évaluer la robustesse des méthodes d'estimation de la masse aux erreurs sur ces entrées.

On crée plusieurs jeux de trajectoires permettant de ne bruiter qu'une variable, plus ou moins intensément. Un jeu de trajectoires est caractérisé par la grandeur bruitée X et l'écart-type de loi normale utilisée σ. Pour obtenir le jeu de trajectoires caractérisé par (X, σ), on part du jeu de trajectoires simulées sans bruit, puis à chaque point de chaque trajectoire on ajoute à la grandeur X la valeur issue du tirage d'une loi normale d'écart-type σ.

Ainsi, si l'on veut tester n valeurs d'écarts-types de bruit pour chacune des cinq grandeurs, on va devoir considérer $5n$ jeux de 1 000 trajectoires chacun.

5.5.2 Résultats

À partir d'un jeu de trajectoires, on calcule une valeur que l'on espère représentative des performances de la méthode d'estimation de la masse. Pour chaque trajectoire du jeu on calcule l'erreur relative $100 \times \frac{\hat{m}_{11} - m_{11}}{m_{11}}$. Ensuite, on calcule la RMSE de ces 1000 erreurs relatives ce qui permet de résumer ces résultats en une seule valeur. Les RMSE obtenues sur

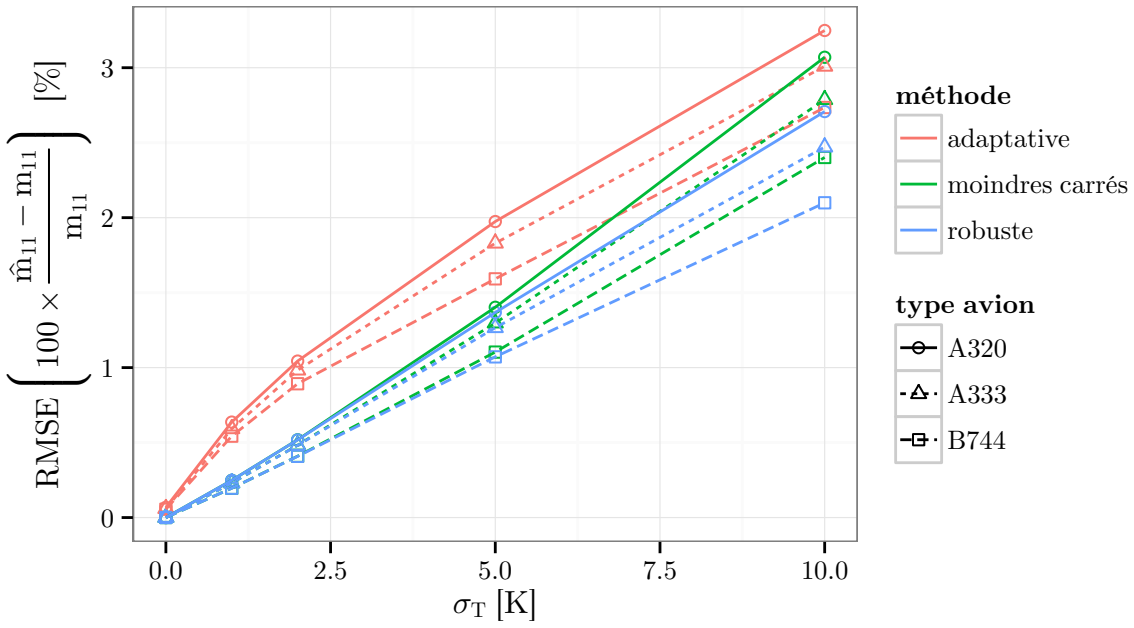


FIGURE 5.5 – Cette figure présente la sensibilité des méthodes d'estimation de la masse aux erreurs sur la température T.

les différents jeux de trajectoires sont représentées sur les figures 5.5 à 5.9. Chaque figure regroupe les résultats obtenus lorsque l'on bruite une grandeur avec différents écarts-types.

Concernant les figures 5.6 et 5.9, les valeurs d'écarts-types choisies sont inspirées des pires cas décrits dans le document Eurocontrol [Eurocontrol 97]. Ce document caractérise les erreurs en position et en vitesse de différents scénarios avec par exemple un nombre variable de radars primaires et secondaires.

Sur les figures 5.5 à 5.9, on voit que toutes les méthodes estiment la masse avec une bonne précision. Dans tous les cas, la précision diminue quand le bruit augmente. Sur ces figures la RMSE ne dépasse pas les 5 % ce qui suggère que les méthodes d'estimation de la masse sont relativement robustes aux erreurs sur la température, l'altitude, la vitesse, l'accélération et le taux de montée.

Étonnamment, l'estimation de la masse est peu sensible aux erreurs sur l'altitude H_p comme illustré par la figure 5.6. Ceci est particulièrement vrai pour la méthode des moindres carrés et la méthode robuste.

Les erreurs sur la vitesse V_a, l'accélération $\frac{dV_a}{dt}$ et le taux de montée $\frac{dH_p}{dt}$ sont celles qui ont le plus d'impact comme on peut le voir sur les figures 5.8 à 5.9. Les plus grandes erreurs d'estimation sont observées en bruitant le taux de montée. Ceci n'est toutefois pas forcément significatif car les grandeurs ne sont pas bruitées avec la même intensité. En effet, il est difficile de comparer des écarts-types de bruit sur des grandeurs différentes. Par exemple, un écart-type de 0,2 $kts.s^{-1}$ pour l'accélération est une grande valeur : les accélérations $\frac{dV_a}{dt}$ sans bruit pour nos données simulées sont comprises entre $-0{,}08$ $kts.s^{-1}$

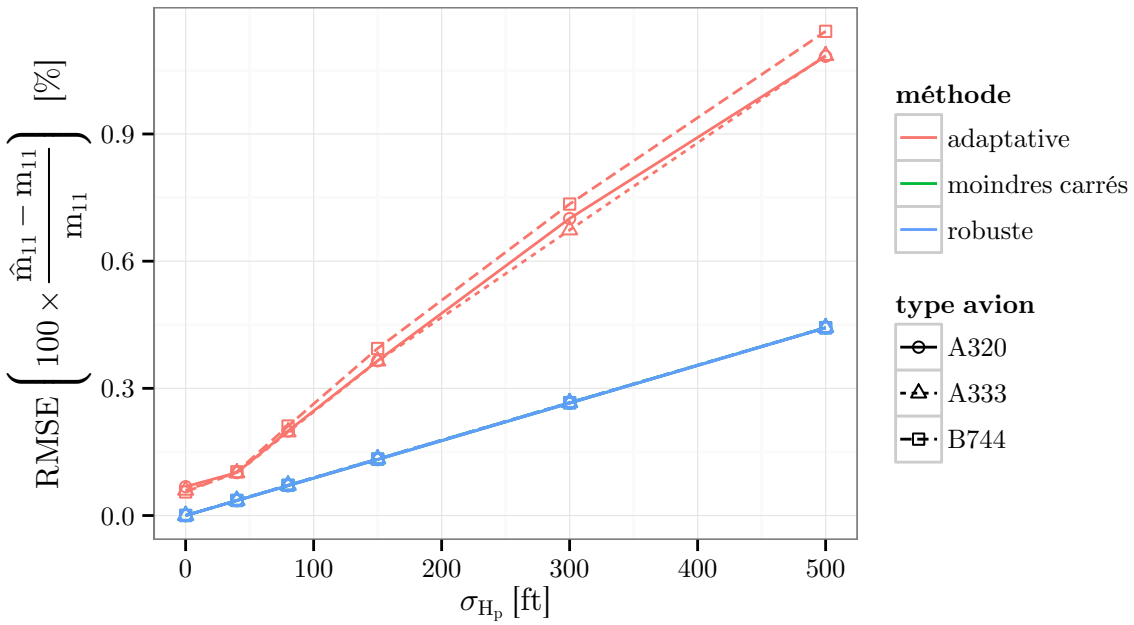


FIGURE 5.6 – Cette figure présente la sensibilité des méthodes d'estimation de la masse aux erreurs sur l'altitude H_p.

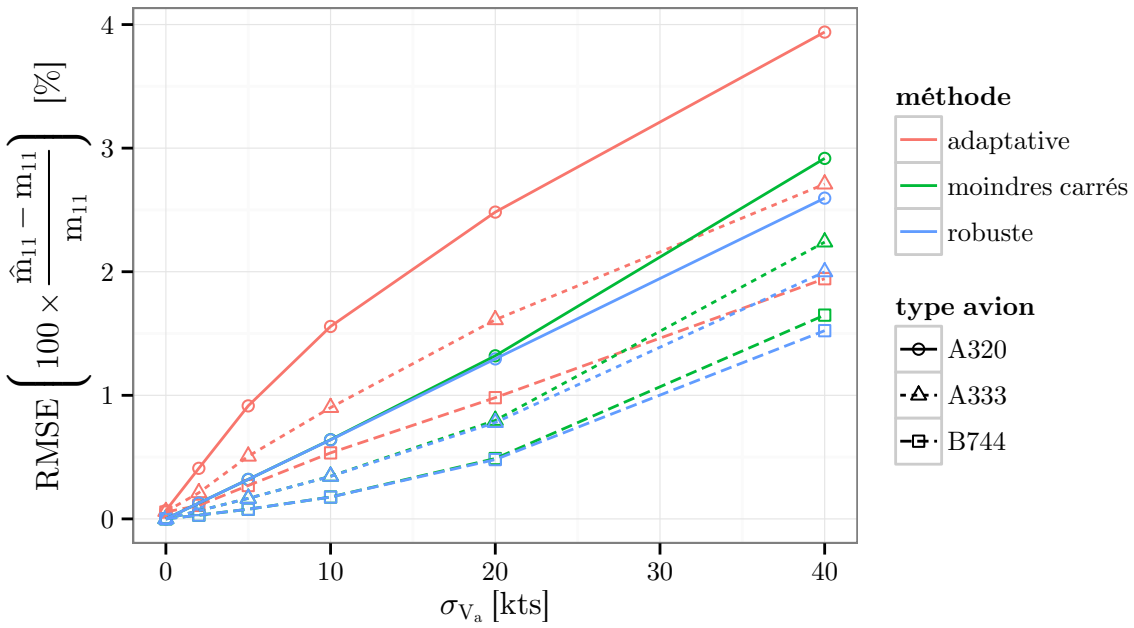


FIGURE 5.7 – Cette figure présente la sensibilité des méthodes d'estimation de la masse aux erreurs sur la vitesse V_a.

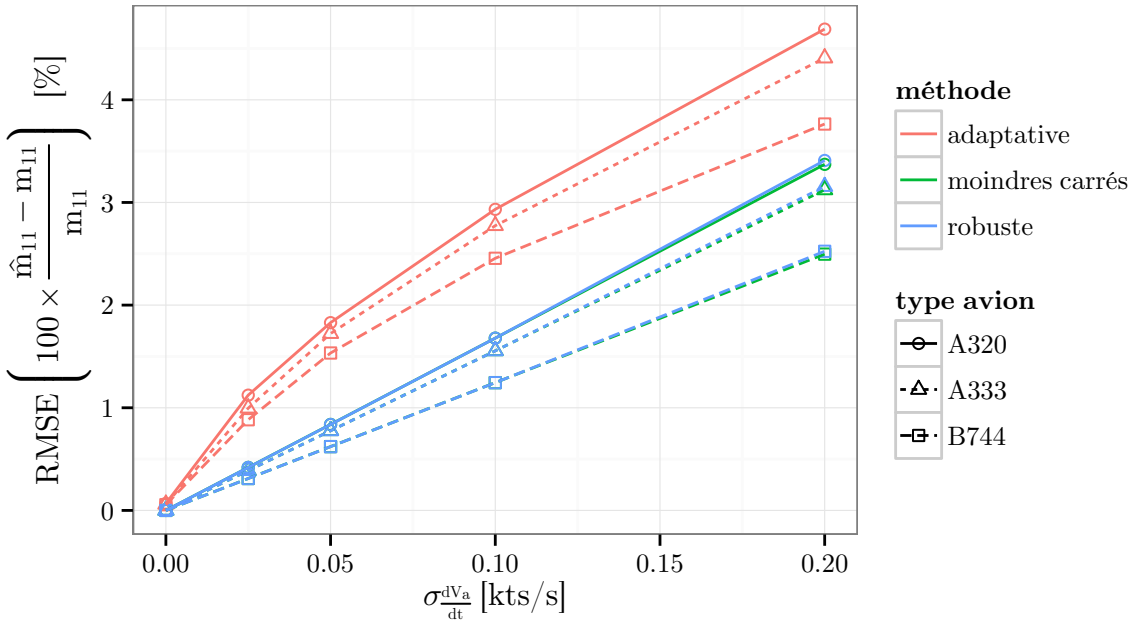


FIGURE 5.8 – Cette figure présente la sensibilité des méthodes d'estimation de la masse aux erreurs sur l'accélération $\frac{dV_a}{dt}$.

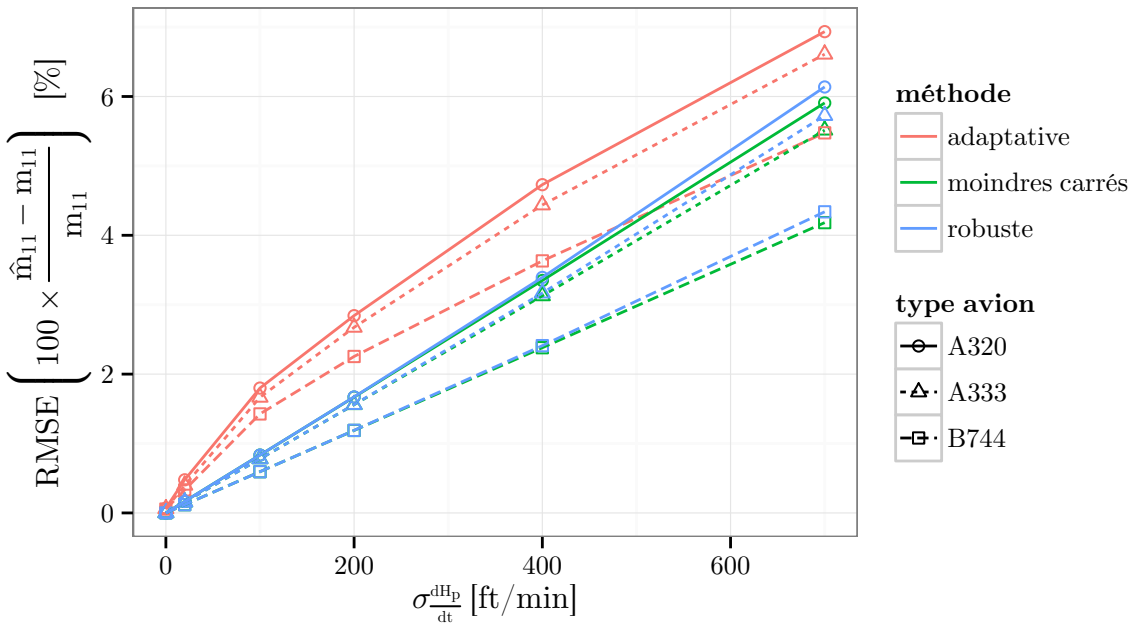


FIGURE 5.9 – Cette figure présente la sensibilité des méthodes d'estimation de la masse aux erreurs sur le taux de montée $\frac{dH_p}{dt}$.

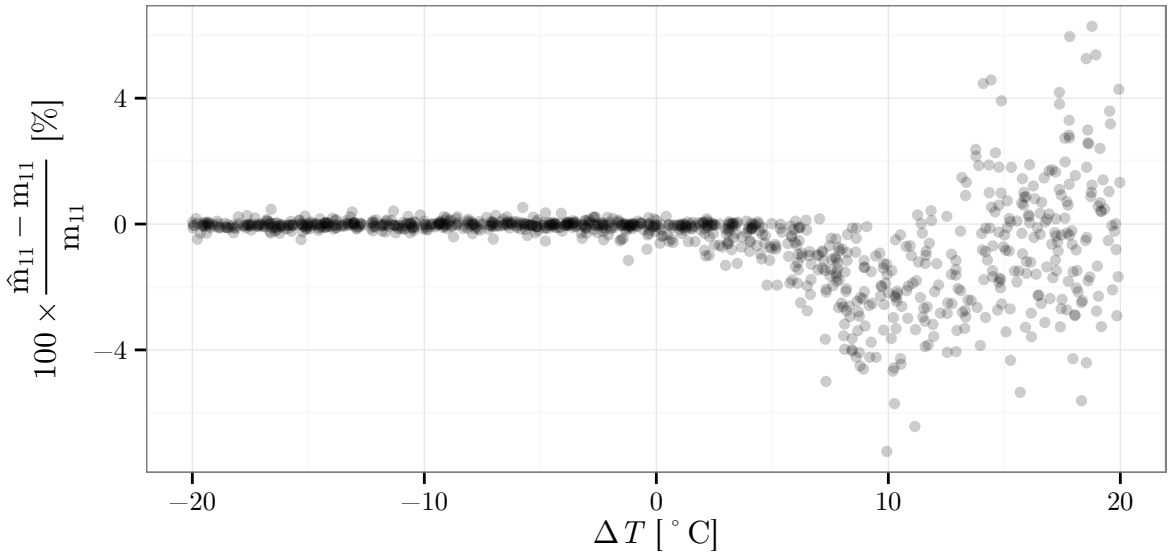


FIGURE 5.10 – Cette figure présente l'erreur relative commise sur la masse en fonction de ΔT. Sur le jeu de trajectoires d'A320 utilisé, la température observée est bruitée avec un bruit gaussien d'écart-type $\sigma_T = 5$ K. $\hat{m}_{11}$ est la masse estimée au onzième point en appliquant la méthode des moindres carrés aux onze premiers points.

et 0,44 $kts.s^{-1}$ pour les trois types d'avions considérés.

Pour chaque méthode, la précision varie de façon similaire pour les différents types avions. Toutefois, on observe une plus grande précision pour le B744 que pour l'A320.

Dans toutes les figures, les méthodes moindres carrés et robuste sont plus précises que la méthode adaptative. Toutefois, les résultats sont plus resserrés si l'on considère un bruit sur la température. Les méthodes moindres carrés et robuste ont une précision similaire. Pour la vitesse V_a, l'accélération $\frac{dV_a}{dt}$, et le taux de montée $\frac{dH_p}{dt}$, la RMSE obtenue avec les méthodes de moindres carrés et robuste est réduite de 20% à 50% par rapport à l'utilisation de la méthode adaptative. Pour l'altitude H_p, cette réduction va jusqu'à 70%.

Au final, avec le nombre de points et les paramètres choisis, les méthodes de moindres carrés et robuste sont plus précises que la méthode adaptative avec les bruits expérimentés. Toutefois, pour toutes les méthodes, les erreurs ne sont que de quelques pourcents.

5.5.3 Influence de ΔT sur l'erreur

Les trajectoires sont générées avec des valeurs aléatoires pour le tuple $(CAS, Mach, \Delta T, mass)$. Pour un bruit fixé, ceci permet de regarder comment est distribuée l'erreur d'estimation en fonction du tuple. En particulier, pour un même bruit sur la température T, on observe que l'erreur d'estimation est plus grande quand ΔT dépasse une valeur seuil.

Pour illustrer ceci, on considère un jeu de trajectoires avec un bruit gaussien d'écart-type $\sigma_T = 5$ K sur la température observée. La masse est estimée sur chaque trajectoire

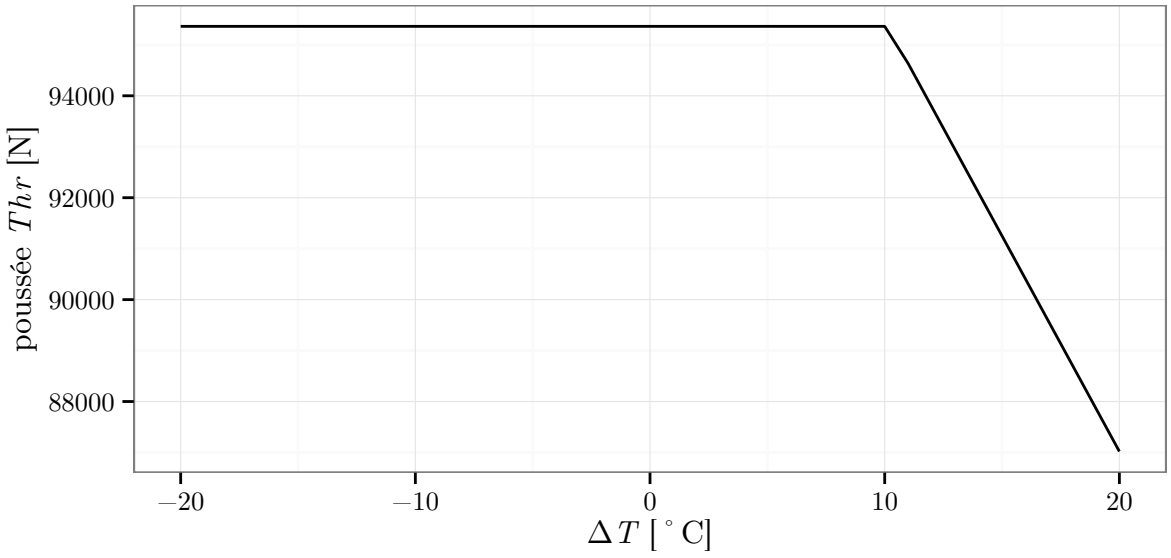


FIGURE 5.11 – Cette figure présente la poussée obtenue par un A320 à $H_p = 18\ 000$ ft en fonction de ΔT.

$(CAS, Mach, \Delta T, mass)$. La figure 5.10 représente les erreurs d'estimation de la masse en fonction de la valeur ΔT avec laquelle on a généré la trajectoire sans bruit. Cette estimation est faite en utilisant la méthode des moindres carrés. Des résultats similaires sont observés avec les autres méthodes. Sur cette figure, l'estimation de la masse est très peu sensible au bruit pour des ΔT inférieurs à un certain seuil. Au-delà de ce seuil, la masse est d'abord sous-estimée puis quand ΔT est suffisamment loin du seuil l'estimation semble sans biais mais avec un plus grand écart-type.

Ceci peut s'expliquer en regardant de plus près le modèle BADA : quand ΔT est supérieur au seuil $C_{Tc,4}$ [2], la poussée *max climb* diminue drastiquement quand la température augmente. La figure 5.11 présente la poussée obtenue par un A320 à $H_p = 18\ 000$ ft en fonction de ΔT. Si on considère une trajectoire avec ΔT proche de ce seuil, quand la température observée est supérieure à la température réelle, la poussée calculée est inférieure à la poussée réelle ce qui conduit à estimer une masse plus faible. Inversement, si la température observée est inférieure à la température réelle, la poussée calculée est similaire à la poussée réelle. Elle est même identique dans le cas où ΔT est inférieure au seuil. Sur les onze points pris en compte, la température observée est tantôt inférieure, tantôt supérieure à la température réelle. Ceci conduit à sous-estimer la masse.

2. Cette constante stockée dans les fichiers BADA est propre à chaque type avion.

5.5.4 Discussion sur les méthodes

Comme vu dans les sous-sections précédentes, les trois méthodes testées sont relativement robustes au bruit introduit sur la trajectoire observée. Les méthodes robuste et des moindres carrées sont toutefois légèrement plus précises dans nos expérimentations.

Au-delà du critère de précision, ces méthodes ont des caractéristiques différentes. La méthode adaptative proposée par Schultz *et. al.* ne repose pas sur un modèle des forces spécifique, un modèle « boîte noire » de la puissance peut être utilisé avec cette méthode. La méthode des moindres carrés et la méthode robuste tirent avantage de l'expression analytique de la modélisation de la puissance. Ces deux méthodes sont fortement liées au modèle utilisé. Par exemple, la méthode des moindres carrés transforme le problème de minimisation de la somme des erreurs au carré en une simple recherche des racines d'un polynôme. Cette transformation exploite la forme analytique du modèle de la puissance. Toutefois, d'autres modèles que BADA ont une forme similaire qui semblent compatible avec la méthode des moindres carrés. C'est le cas du *Enhanced Jet Performance Model* ([Kaiser 11]).

Dans les trois méthodes, la masse est ajustée pour que l'excès de puissance spécifique corresponde à la variation d'énergie spécifique en supposant connue la loi de poussée choisie par le pilote. Toutefois, les équations (5.1) résultant de cette égalité sont utilisées différemment dans les méthodes.

Dans la méthode adaptative, la masse est estimée en considérant les points de manière séquentielle. À chaque nouveau point, on utilise l'estimation précédente et l'équation (5.1) pour calculer une nouvelle estimation. La différence entre la nouvelle estimation et l'ancienne est bornée à 2,5 % de la masse de référence. Sans ce mécanisme, cette méthode est très peu précise. Toutefois, il faut plusieurs points pour que la masse estimée puisse atteindre n'importe quelle valeur dans l'intervalle $mass_{ref} \times [0.8; 1.2]$. Cette limitation mise à part, la variation de la masse estimée n'est pas contrainte entre les points. Dans la méthode des moindres carrés et la méthode robuste, la variation de la masse estimée est contrainte par la modélisation BADA de la consommation du carburant.

5.5.5 Limites et conclusion de cette étude

Cette étude repose sur des données simulées. Ceci permet une connaissance parfaite du modèle physique sous-jacent. De plus, on connaît la vraie masse et les vraies valeurs des grandeurs T, V_a, H_p, $\frac{dH_p}{dt}$ et $\frac{dV_a}{dt}$. Le but de cette étude n'est pas de prédire les performances des méthodes sur des données réelles mais de comparer leur robustesse lorsque les grandeurs observées sont bruitées.

Concernant le jeu de trajectoires, la loi uniforme sur le tuple $(CAS, Mach, \Delta T, mass)$ utilisé ne correspond pas au trafic réel. Toutefois, elle permet d'avoir une grande variété de trajectoires. De même, le bruit gaussien ajouté sur chaque observation d'une des grandeurs n'est pas réaliste. Il y a par exemple des erreurs systématiques dans la mesure de la position et de la vitesse ([Eurocontrol 97]).

Sur nos jeux de trajectoires simulées, l'erreur relative est faible au vu des valeurs d'écart-

type testées. Cette étude montre que la précision de l'estimation est variable suivant les trajectoires. En effet, l'écart-type de cette erreur relative est plus grand lorsque l'atmosphère est suffisamment chaude pour que ΔT dépasse un certain seuil. Ce seuil correspond à un seuil dans la modélisation de la poussée BADA. Les performances des méthodes sont similaires avec toutefois une plus grande précision pour la méthode des moindres carrés et la méthode robuste.

5.6 Comparaison sur des données réelles

Dans cette sous-section, on utilise les jeux de trajectoires issues de la section 4.1. Ceci va permettre de comparer les résultats obtenus avec les méthodes d'estimation de la masse présentées dans ce chapitre et les méthodes d'apprentissage présentées dans le chapitre 6.5. Pour rappel, le jeu d'exemples est uniquement composé de trajectoires d'avion de type A320 en phase de montée. Chaque trajectoire est constituée d'une suite de 51 points espacés de $\Delta t = 15$ s. Les exemples sont construits de manière à avoir le 11^e point à une altitude H_p de 18 000 ft. Les 11 premiers points sont ceux connus. Les 40 points suivants sont ceux à prédire. On a ainsi un horizon de prédiction de 10 min.

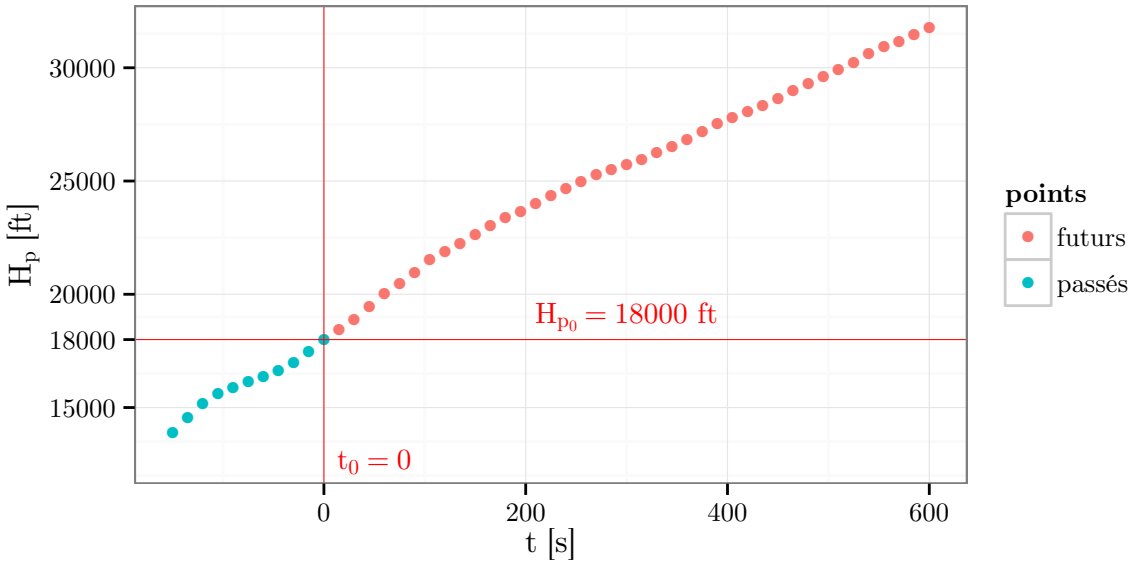


FIGURE 5.12 – Cette figure représente une trajectoire de notre jeu d'exemples.

La comparaison des méthodes d'estimation de la masse sur des données réelles est délicate. Sur nos données réelles, la masse n'est pas connue, il est donc *a priori* impossible de déterminer l'écart entre la masse estimée et la masse réelle.

Toutefois, il est possible de comparer les performances obtenues en utilisant la masse estimée pour prédire la trajectoire future. De plus, on va aussi pouvoir appliquer nos

méthodes d'estimation de masse sur les points futurs.

5.6.1 Comparaison de la prédiction

La qualité de la prédiction des points futurs obtenue en utilisant la masse estimée peut être évaluée. Pour utiliser le modèle BADA, il faut avoir l'état initial $(m, V_a, \Psi_i, H_p, x, y, t)_{t_0}$, un modèle météo, le profil de poussée et le profil de vitesse air. On souhaite prédire la trajectoire future à partir de la date t_0. À cette date, tout l'état initial est connu sauf la masse. Comme on s'intéresse à l'impact des méthodes d'estimation de la masse sur la prédiction de trajectoire, on va supposer que l'on connaît le profil de poussée et le profil de vitesse air choisis par le pilote. La commande de poussée choisie par le pilote est supposée être *max climb*. Pour le profil de vitesse air, on prend $V_a = V_{a_{cible}}(t)$, avec $V_{a_{cible}}$ qui est issue de l'interpolation linéaire des vitesses observées aux points futurs. Ainsi, pour chaque trajectoire, les vitesses calculées à partir de BADA et celles observées sur la trajectoire future seront identiques. Au final, seule la masse est manquante et suivant sa valeur, les altitudes prédites différeront de celles observées.

Les tableaux 5.2 et 5.3 présentent des statistiques calculées sur l'ensemble des valeurs $\left(H_p^{(pred)}(\hat{m}_{11}) - H_p^{(obs)}\right)(t = 600 \text{ s})$, avec $\hat{m}_{11}$ la masse estimée sur les points passés. En plus des méthodes d'estimation de masse, les tableaux contiennent les résultats obtenus en utilisant la masse de référence $mass_{ref}$. Ceci permet d'évaluer le gain obtenu avec les méthodes d'estimation de la masse par rapport à l'emploi de la masse de référence. En l'occurrence, la réduction de la RMSE va de 40 % à 50 % en fonction de la méthode utilisée. Du point de vue de la RMSE, les quatre méthodes ont des performances assez similaires. Les méthodes robuste et adaptative sont celles donnant les meilleurs résultats.

Plus précisément, en regardant la moyenne, on s'aperçoit que l'altitude finale est sous-estimée ce qui suggère que la masse est surestimée si l'on exclut des erreurs sur le modèle physique ou la loi de poussée. Des quatre méthodes, la méthode adaptative est celle surestimant le moins la masse. Concernant l'écart-type, la méthode robuste est celle présentant la plus faible valeur tandis que les valeurs des autres méthodes sont comparables. La combinaison du biais et de l'écart-type fait que les méthodes robuste et adaptative ont une RMSE plus faible que les autres.

TABLE 5.2 – Statistiques sur l'écart entre les trajectoires prédites et le dernier plot observés pour les trajectoires Mode-C. Ces statistiques sont calculées sur l'ensemble des valeurs $\left(H_p^{(pred)}(\hat{m}_{11}) - H_p^{(obs)}\right)(t = 600 \text{ s})$. Chacune de ces valeurs est associée à une trajectoire. Les trajectoires sont calculées avec $V_a(t) = V_a{}^{(obs)}(t)$.

méthode	moy [ft]	écart-type [ft]	moy abs [ft]	rmse [ft]	max abs [ft]
$mass_{ref}$	-82.4	1479	1168	1481	5493
naïve	-524	651	628	836	6874
adaptative	-303	685	582	749	5535
moindres carrés	-532	653	631	843	6033
robuste	-503	595	603	779	6843

TABLE 5.3 – Statistiques sur l'écart entre les trajectoires prédites et les plots observés pour les trajectoires Mode-S. Ces statistiques sont calculées sur l'ensemble des valeurs $\left(H_p^{(pred)}(\hat{m}_{11}) - H_p^{(obs)}\right)(t = 600 \text{ s})$. Chacune de ces valeurs est associée à une trajectoire. Les trajectoires sont calculées avec $V_a(t) = V_a{}^{(obs)}(t)$.

méthode	moy [ft]	écart-type [ft]	moy abs [ft]	rmse [ft]	max abs [ft]
$mass_{ref}$	-810	1133	1079	1392	5531
naïve	-598	602	673	849	6574
adaptative	-551	554	651	781	4171
moindres carrés	-584	606	663	842	7150
robuste	-571	550	645	792	6371

Même en estimant la masse sur les points passés, les erreurs maximum restent grandes, plus de 5 500 ft. En effet, on a ajusté la masse sur les points passés. Ainsi, la puissance calculée et observée vont correspondre sur ces points passés. Lorsque l'on calculera la prédiction avec la masse estimée, si la puissance est mal modélisée alors la puissance calculée va dévier de celle observée lors de la montée. Cette mauvaise modélisation de la puissance peut avoir plusieurs sources. Elle peut venir d'une mauvaise modélisation des forces avec un modèle surestimant la traînée pour les altitudes élevées par exemple. Elle peut aussi venir d'une mauvaise hypothèse sur la loi de poussée utilisée. Cette dernière cause possible est la plus probable pour des erreurs grandes. En effet, pour ces cas, la puissance observée varie brutalement. Ces cas sont illustrés par les figures 5.13 et 5.14. Pour la figure 5.13, l'avion n'a probablement pas une commande de poussée constante sur les points passés ce qui engendre une sur-estimation de la masse et une erreur de 6 033 ft à $t = 600$ s. C'est la trajectoire Mode-C sur laquelle on a le plus d'erreur avec la méthode des moindres carrés. Pour la figure 5.14, la puissance observée baisse brutalement à $t = 0$. Ainsi, la trajectoire prédite à $t = 600$ s surestime l'altitude de 2 669 ft. C'est la trajectoire Mode-S qui surestime le plus l'altitude à $t = 600$ s en estimant la masse avec la méthode des moindres carrés.

5.6.2 Distribution de la masse estimée sur les points passés

Les figures 5.15 et 5.16 présentent les histogrammes des masses estimées obtenues avec les différentes méthodes. Les distributions obtenues avec les différentes méthodes sont assez similaires. Toutes les masses estimées par la méthode adaptative sont par « construction » comprises entre 51 200 kg et 76 800 kg. Les masses estimées par les autres méthodes n'ont pas cette contrainte, il y a des masses estimées qui dépassent la masse maximale de 77 000 kg enregistrée dans les fichiers BADA pour le type A320. Une masse estimée si élevée n'est pas réaliste. Si l'on écarte l'erreur de modélisation de la puissance, ceci suggère que l'hypothèse de poussée *max climb* n'est pas vraie pour tous les points passés de la trajectoire.

La moyenne des masses estimées des trajectoires Mode-C est supérieure de 2 000 kg à celle des données Mode-S. Les masses estimées sur les trajectoires Mode-S et Mode-C forment des distributions unimodales. Le mode associé aux trajectoires Mode-S est plus

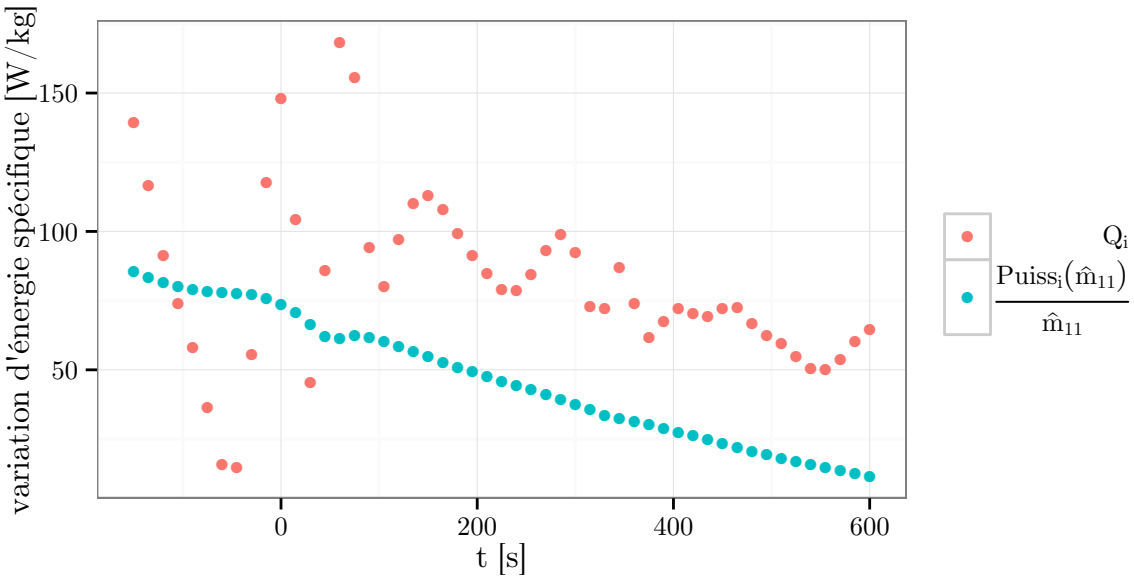


FIGURE 5.13 – Cette figure représente la puissance $\frac{Puiss_i(\hat{m}_{11})}{\hat{m}_{11}}$ calculée et la puissance observée Q_i. De par les équations (5.1), ces deux grandeurs doivent être égales. $\hat{m}_{11}$ est la masse estimée à $t = 0$ sur les onze premiers points de la trajectoire. C'est la trajectoire Mode-C sur laquelle on a le plus d'erreur avec la méthodes des moindres carrés.

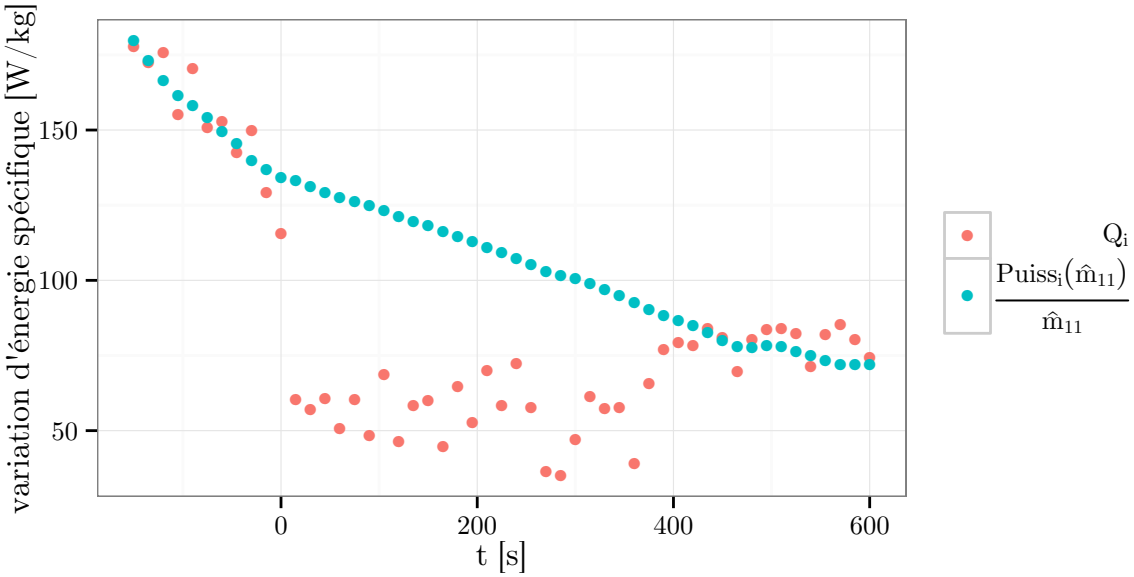


FIGURE 5.14 – Cette figure représente la puissance $\frac{Puiss_i(\hat{m}_{11})}{\hat{m}_{11}}$ calculée et la puissance observée Q_i. De par les équations (5.1), ces deux grandeurs doivent être égales. $\hat{m}_{11}$ est la masse estimée à $t = 0$ sur les onze premiers points de la trajectoire. C'est la trajectoire Mode-S qui surestime le plus l'altitude à $t = 600$ s en estimant la masse avec la méthode des moindres carrés.

élevé que le mode associé aux trajectoires Mode-C. Les masses estimées sur les trajectoires Mode-S sont plus resserrées que celles estimées sur les trajectoires Mode-C.

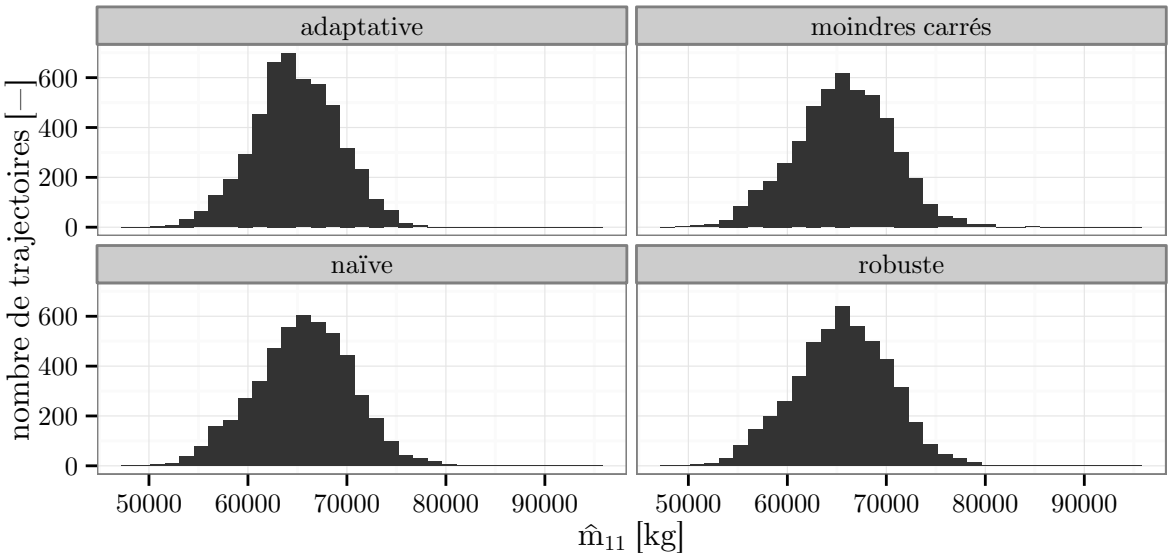


FIGURE 5.15 – Cette figure présente un histogramme de la masse estimée sur les trajectoires Mode-C en utilisant les 11 premiers points.

5.6.3 Discussion sur la masse estimée sur les points futurs

Sur nos données réelles, la masse n'est pas connue, il est donc *a priori* impossible de déterminer l'écart entre la masse estimée et la masse réelle. On peut toutefois estimer une masse sur les 41 points futurs de la trajectoire. Cette masse $\hat{m}_{11,futur}$ est estimée avec la méthode des moindres carrés en utilisant le modèle qui servira aussi à prédire la trajectoire future. Ainsi la trajectoire calculée à partir de cette masse sera très proche de la trajectoire observée. Les tableaux 5.4 et 5.5 quantifient cet écart. L'écart sur les altitudes est drastiquement réduit comparée à l'utilisation de $\hat{m}_{11}$. Mais cet écart existe toujours. Comme évoqué précédemment, cela est inhérent aux erreurs sur les grandeurs observées, aux erreurs de modélisation de la puissance et aux hypothèses sur la loi de poussée utilisée. Même si les résultats en prédiction sont bons, la masse estimée $\hat{m}_{11,futur}$ peut être assez différente de la masse réelle de l'avion.

Toutefois, la masse estimée $\hat{m}_{11,futur}$ est croissante avec la distance à parcourir comme illustré par la figure 5.17. C'est une propriété attendue concernant la masse réelle. On la retrouve également pour nos masses estimées $\hat{m}_{11,futur}$, et ce malgré les sources d'erreurs nuisant à l'estimation de la masse réelle de l'avion.

Les figures 5.18 et 5.19 présentent la masse $\hat{m}_{11,futur}$ estimée sur les points futurs en fonction de la masse $\hat{m}_{11}$ estimée sur les points passés. La droite rouge d'équation $y = x$

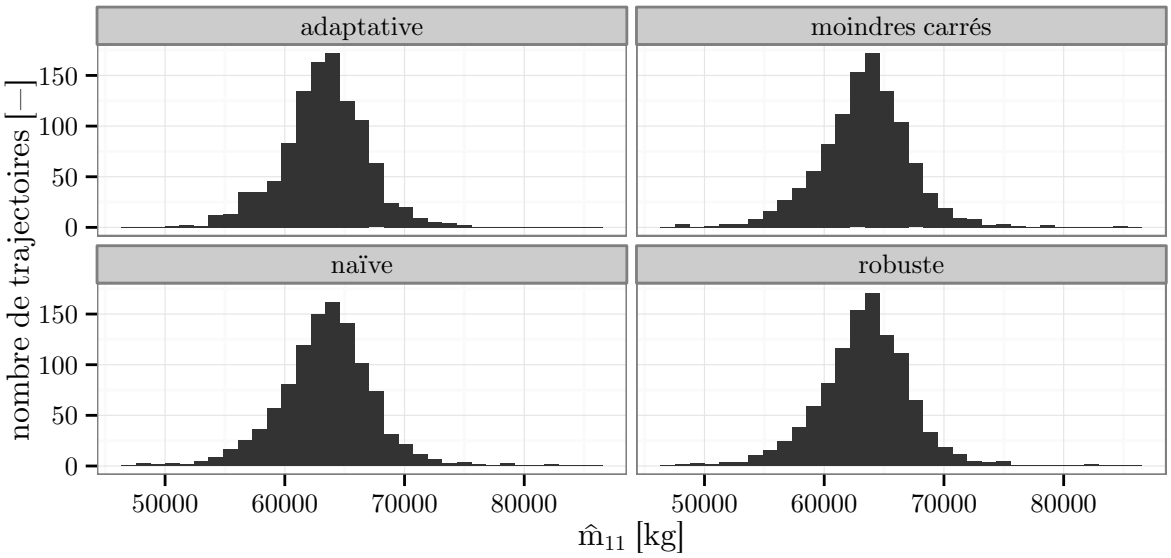


FIGURE 5.16 – Cette figure présente un histogramme de la masse estimée sur les trajectoires Mode-S en utilisant les 11 premiers points.

TABLE 5.4 – Statistiques sur l'écart entre les trajectoires prédites et le dernier plot observés pour les trajectoires Mode-C et Mode-S. Ces statistiques sont calculées sur l'ensemble des valeurs $\left(H_p^{(pred)}(\hat{m}_{11,futur}) - H_p^{(obs)}\right)(t \geqslant 0)$. Chacune de ces valeurs est associée à une trajectoire. Les trajectoires sont calculées avec $V_a(t) = {V_a}^{(obs)}(t)$.

trajectoires	moy [ft]	écart-type [ft]	moy abs [ft]	rmse [ft]	max abs [ft]
Mode-C	31.7	144	111	147	1407
Mode-S	81.8	185	143	202	2245

permet de lire facilement les écarts entre ces deux masses. Avec ces figures, on confirme le fait que la majorité des masses $\hat{m}_{11}$ sous-estiment la valeur nécessaire à une bonne prédiction de trajectoire. Les tableaux 5.6 et 5.7 quantifient cet écart entre les masses $\hat{m}_{11}$ et $\hat{m}_{11,futur}$.

Avec l'hypothèse *max climb*, la masse minimisant l'écart entre la puissance observée et celle calculée peut très bien être supérieure à la masse maximale de 77 000 kg du modèle BADA. On a d'ailleurs vu que les masses estimées $\hat{m}_{11}$ avec les méthodes autres qu'adaptative peuvent très bien dépasser la masse maximale. Ceci est observé aussi bien pour les données Mode-C que Mode-S. Concernant la masse $\hat{m}_{11,futur}$, le constat est différent. Pour les trajectoires Mode-C, la plage des masses $\hat{m}_{11,futur}$ est plus restreinte que celle des masses estimées $\hat{m}_{11}$. Les valeurs élevées de $\hat{m}_{11}$ surestiment les masses $\hat{m}_{11,futur}$. Pour les trajectoires Mode-S, la plage des masses $\hat{m}_{11,futur}$ est assez similaire à celle des masses estimées $\hat{m}_{11}$. En particulier, on observe qu'une masse $\hat{m}_{11}$ estimée à 100 000 kg est confirmée par

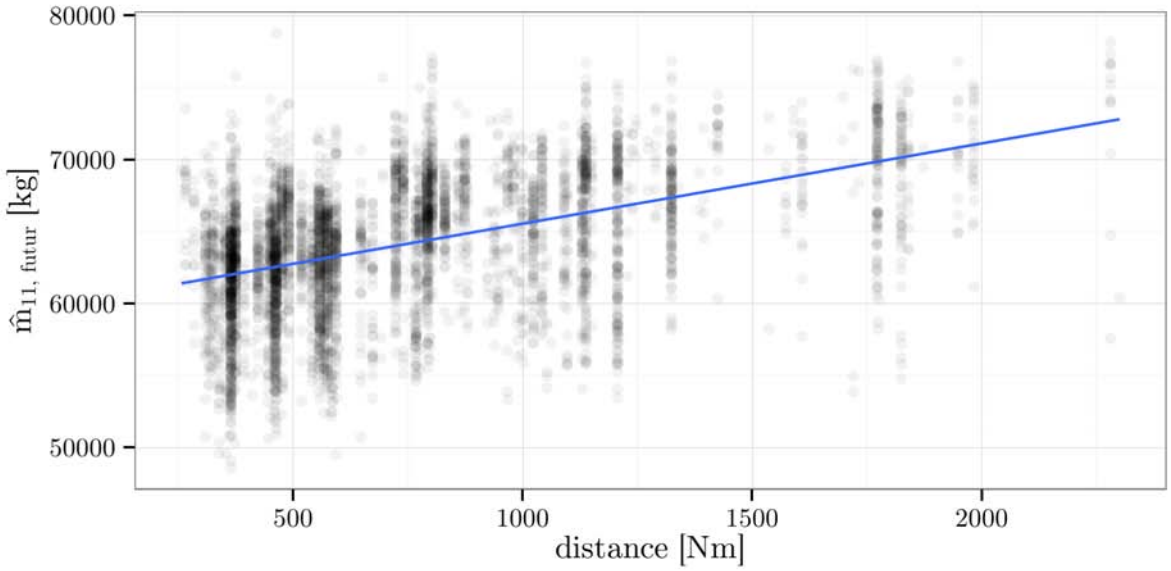


FIGURE 5.17 – Cette figure présente la masse $\hat{m}_{11,futur}$ en fonction de la distance séparant les aéroports de départ et d'arrivée. La droite bleue est issue de l'ajustement par moindres carrés d'un modèle linéaire « $y = ax + b$ ». Cette distance est calculée sur une orthodromie à partir des positions des aéroports sur le globe. Cette figure présente uniquement les trajectoires Mode-C. Ces informations n'apparaissent dans notre jeu de trajectoires Mode-S.

TABLE 5.5 – Statistiques sur l'écart entre les trajectoires prédites et le dernier plot observés pour les trajectoires Mode-C et Mode-S. Ces statistiques sont calculées sur l'ensemble des valeurs $\left(H_p^{(pred)}(\hat{m}_{11,futur}) - H_p^{(obs)}\right)(t = 600 \text{ s})$. Chacune de ces valeurs est associée à une trajectoire. Les trajectoires sont calculées avec $V_a(t) = V_a{}^{(obs)}(t)$.

trajectoires	moy [ft]	écart-type [ft]	moy abs [ft]	rmse [ft]	max abs [ft]
Mode-C	-63.1	121	105	137	817
Mode-S	-145	145	165	205	967

TABLE 5.6 – Statistiques, en kilogrammes, sur l'écart entre les trajectoires lissées et le dernier plot observés pour les trajectoires Mode-C. Ces statistiques sont calculées sur l'ensemble des valeurs $\hat{m}_{11} - \hat{m}_{11,futur}$. Chacune de ces valeurs est associée à une trajectoire.

méthode	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	-119	4703	3751	4704	15397
naïve	1524	2050	1868	2554	27281
adaptative	747	2054	1712	2186	15241
moindres carrés	1552	2034	1868	2558	23335
robuste	1452	1825	1773	2332	24505

une masse $\hat{m}_{11,futur}$ tout aussi élevée. Pour cet exemple, la trajectoire prédite avec $\hat{m}_{11}$ sera plus proche de celle observée que celle prédite avec la masse maximale.

5.6.4 Qualité de l'ajustement et qualité de la prédiction

La méthode des moindres carrés estime la masse en ajustant la puissance calculée à celle observée. La qualité de ces ajustements peut nous éclairer quant à la qualité de la prédiction. Tout d'abord, pour quantifier cette qualité, l'équation (5.14) définit $e^{[i:j]}(m)$ qui est la RMSE sur les points i à j des écarts entre la puissance observée et celle calculée avec la masse m. Pour rappel, $\hat{m}_{11}$ est la masse minimisant $e^{[1:11]}$ et $\hat{m}_{11,futur}$ est celle minimisant $e^{[11:51]}$. Les erreurs d'ajustement considérées sont $e^{[1:11]}(\hat{m}_{11})$, $e^{[11:51]}(\hat{m}_{11,futur})$

TABLE 5.7 – Statistiques, en kilogrammes, sur l'écart entre les trajectoires lissées et le dernier plot observés pour les trajectoires Mode-S. Ces statistiques sont calculées sur l'ensemble des valeurs $\hat{m}_{11} - \hat{m}_{11,futur}$. Chacune de ces valeurs est associée à une trajectoire.

méthode	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	1993	3412	3079	3950	14825
naïve	1436	1868	1752	2356	22343
adaptative	1274	1634	1673	2071	16205
moindres carrés	1395	1887	1716	2346	24791
robuste	1348	1680	1657	2153	21517

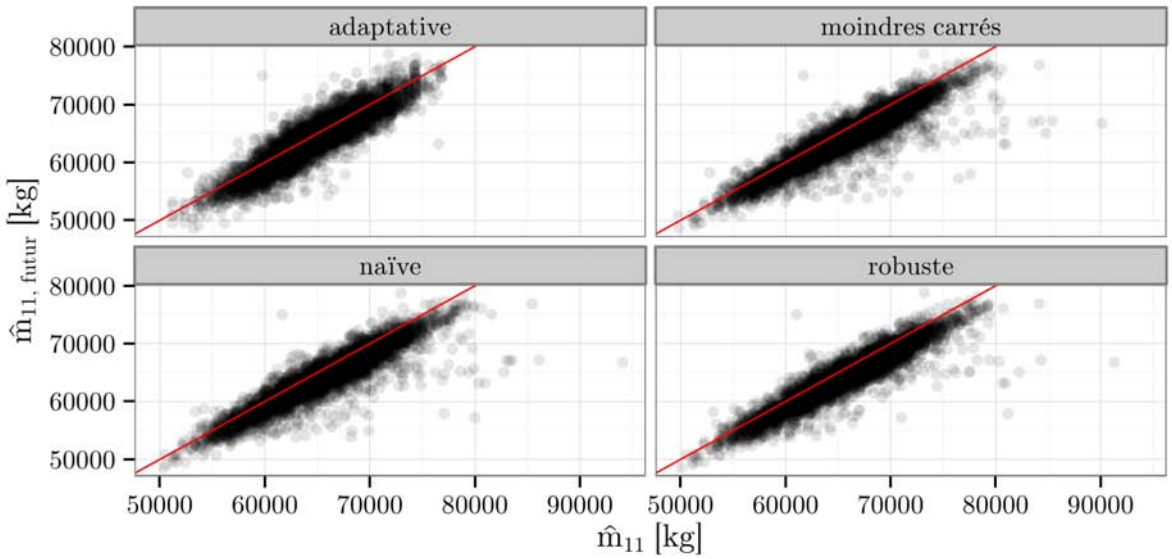


FIGURE 5.18 – Cette figure présente la masse $\hat{m}_{11,futur}$ en fonction de la masse $\hat{m}_{11}$. La droite rouge a pour équation $y = x$. Elle permet de lire les écarts entre les deux masses. Cette figure présente uniquement les trajectoires Mode-C.

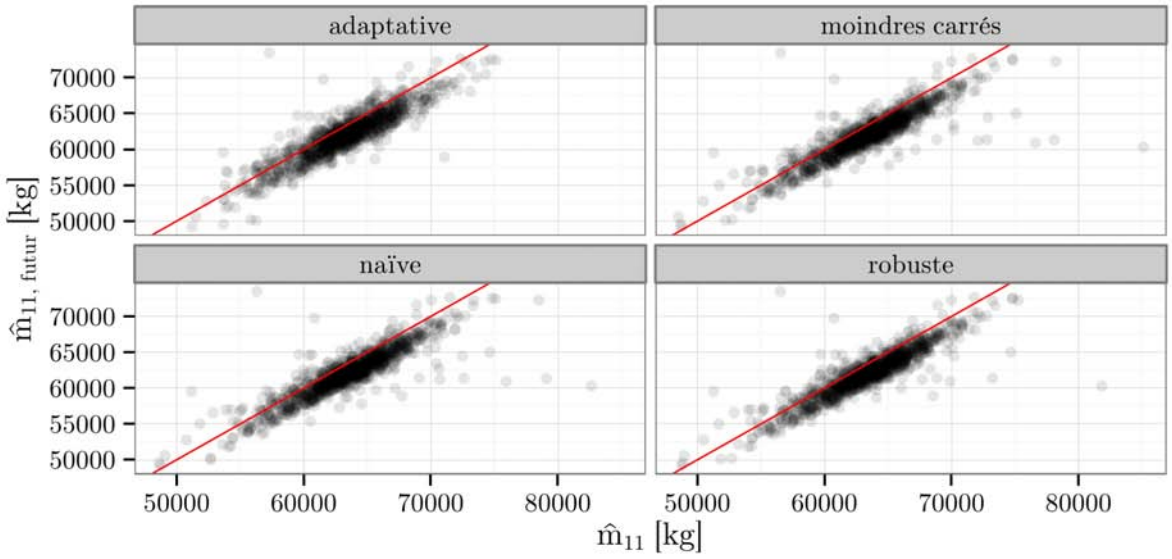


FIGURE 5.19 – Cette figure présente la masse $\hat{m}_{11,futur}$ en fonction de la masse $\hat{m}_{11}$. La droite rouge a pour équation $y = x$. Elle permet de lire les écarts entre les deux masses. Cette figure présente uniquement les trajectoires Mode-S.

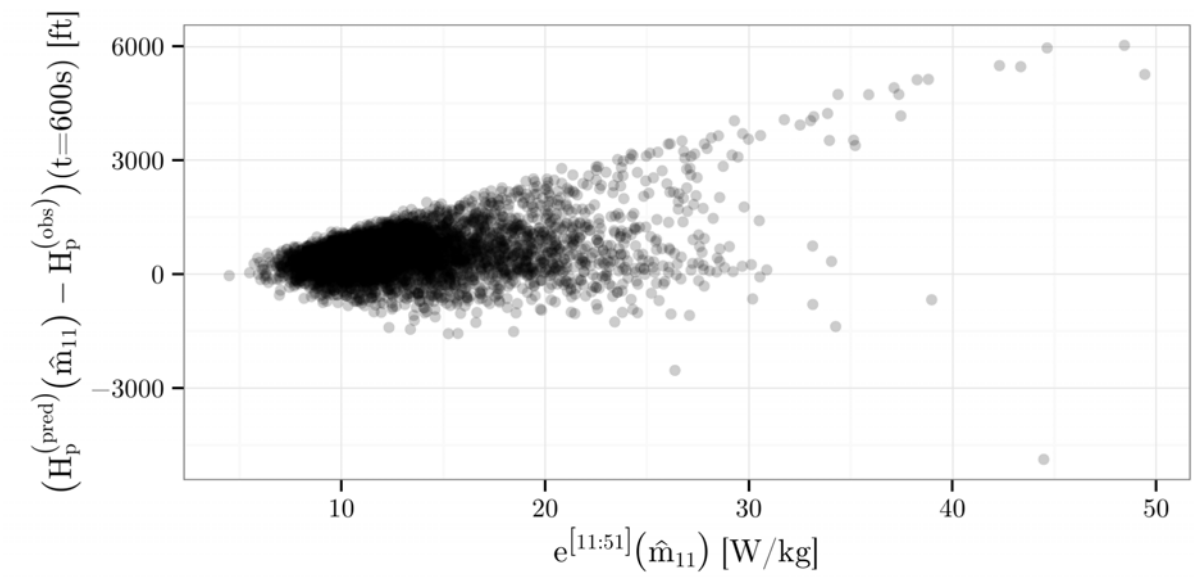


FIGURE 5.20 – Cette figure présente l'erreur sur l'altitude à $t = 600$ s en fonction de l'erreur faite sur l'excès de puissance spécifique sur les points futurs en utilisant la masse estimée $\hat{m}_{11}$. Le jeu de trajectoire Mode-C est utilisé.

et $e^{[11:51]}(\hat{m}_{11})$.

$$e^{[i:j]}(m) = \sqrt{\frac{1}{j-i+1}\mathcal{E}^{[i:j]}{}_{approx}(m)} \tag{5.14}$$

Cette dernière est la RMSE calculée sur les points futurs entre la puissance observée et celle calculée avec la masse $\hat{m}_{11}$ estimée sur les points passés. Une valeur élevée de cette dernière rend possible une grande erreur de prédiction. Les figures 5.20 et 5.21 illustrent cette relation entre l'erreur $e^{[11:51]}(\hat{m}_{11})$ et l'erreur finale sur l'altitude. On remarque qu'une grande erreur $e^{[11:51]}(\hat{m}_{11})$ n'implique pas forcément une grand erreur sur l'altitude. Inversement, une grande erreur sur l'altitude n'est possible que lorsque $e^{[11:51]}(\hat{m}_{11})$ est grand. L'erreur sur la puissance va entraîner une variation d'altitude différente de la trajectoire réelle. Toutefois, ceci ne va pas à l'encontre d'une trajectoire prédite qui oscille autour de la trajectoire réelle tout en restant proche de cette dernière.

On va maintenant s'intéresser à cette erreur $e^{[11:51]}(\hat{m}_{11})$ et sa relation éventuelle avec les erreurs d'ajustements $e^{[1:11]}(\hat{m}_{11})$ et $e^{[11:51]}(\hat{m}_{11,futur})$. Pour chaque trajectoire on a fait deux ajustements, l'un sur les 11 points passés, l'autre sur les 41 points futurs. Les résultats de ces deux ajustements ne sont pas disponibles au même moment et ne donnent pas la même information. L'ajustement passé peut nous renseigner sur la qualité de l'estimation de la masse $\hat{m}_{11}$. Si l'ajustement est mauvais, la puissance observée sur les points passés ne correspond pas à la modélisation qui en est faite. Il y a sûrement eu des variations non modélisées de la commande de poussée. Ainsi, la masse est probablement mal estimée. Un

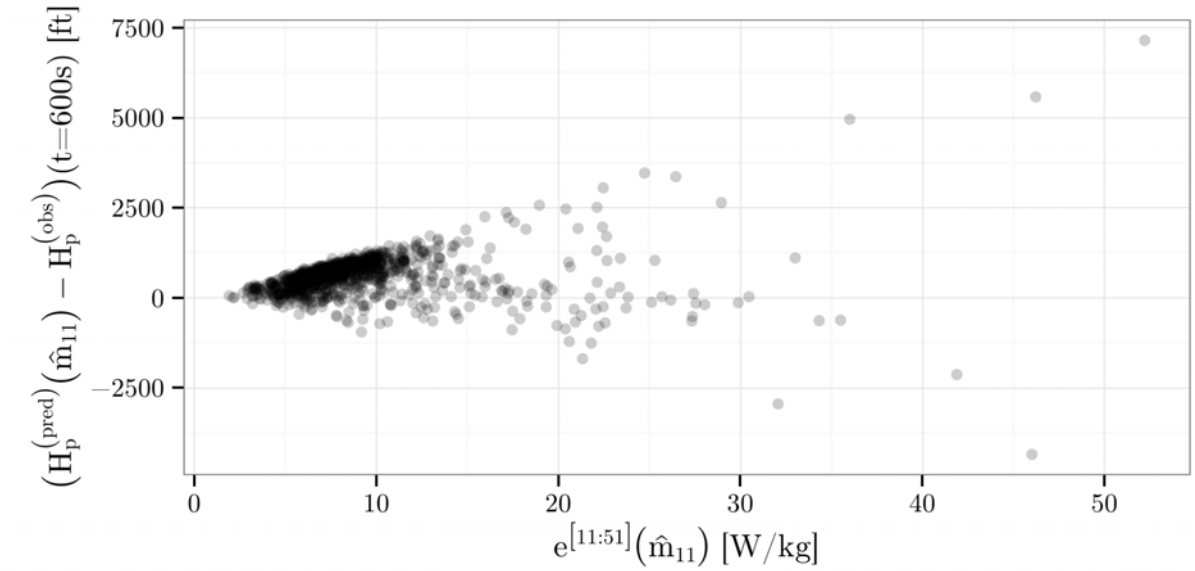


FIGURE 5.21 – Cette figure présente l'erreur sur l'altitude à $t = 600$ s en fonction de l'erreur faite sur l'excès de puissance spécifique sur les points futurs en utilisant la masse estimée $\hat{m}_{11}$. Le jeu de trajectoire Mode-S est utilisé.

exemple de ce cas est présenté figure 5.13. La qualité de l'ajustement sur les points futurs nous renseigne sur la qualité de prédiction que l'on peut obtenir dans le meilleur cas. Si $e^{[11:51]}(\hat{m}_{11,futur})$ est grand, même dans le meilleur cas la prédiction peut être mauvaise. Le meilleur cas étant d'avoir une masse estimée $\hat{m}_{11}$ proche de $\hat{m}_{11,futur}$. Les figures 5.22 et 5.23 illustrent cette analyse.

Ainsi, si $e^{[11:51]}(\hat{m}_{11})$ est élevée alors la prédiction peut être mauvaise. Il est donc intéressant de pouvoir estimer $e^{[11:51]}(\hat{m}_{11})$ avec les informations dont on dispose lors de la prédiction. Les figures 5.22 et 5.23 montrent que si $e^{[1:11]}(\hat{m}_{11})$ est grand alors avoir $e^{[11:51]}(\hat{m}_{11})$ grand est plus probable. Malheureusement, la majorité des grandes valeurs de $e^{[11:51]}(\hat{m}_{11})$ s'observent avec $e^{[1:11]}(\hat{m}_{11})$ petit et $e^{[11:51]}(\hat{m}_{11,futur})$ grand. En effet, ce dernier minore $e^{[11:51]}$.

En conclusion, si $e^{[1:11]}(\hat{m}_{11})$ est grand alors $e^{[11:51]}(\hat{m}_{11})$ est probablement grand ce qui permet d'avoir potentiellement de grandes erreurs. Malheureusement, la réciproque est fausse. On peut avoir $e^{[11:51]}(\hat{m}_{11})$ grand sans que $e^{[1:11]}(\hat{m}_{11})$ soit grand. Ainsi, à partir des points passés, on peut détecter des cas pour lesquels une grande erreur est possible mais on ne les détectera pas tous.

5.6.5 Conclusion de l'étude comparative sur les données réelles

Dans cette étude on a testé les méthodes d'estimation de la masse sur les trajectoires Mode-C et Mode-S. Seul l'A320 est représenté dans ces jeux de trajectoires. En appliquant

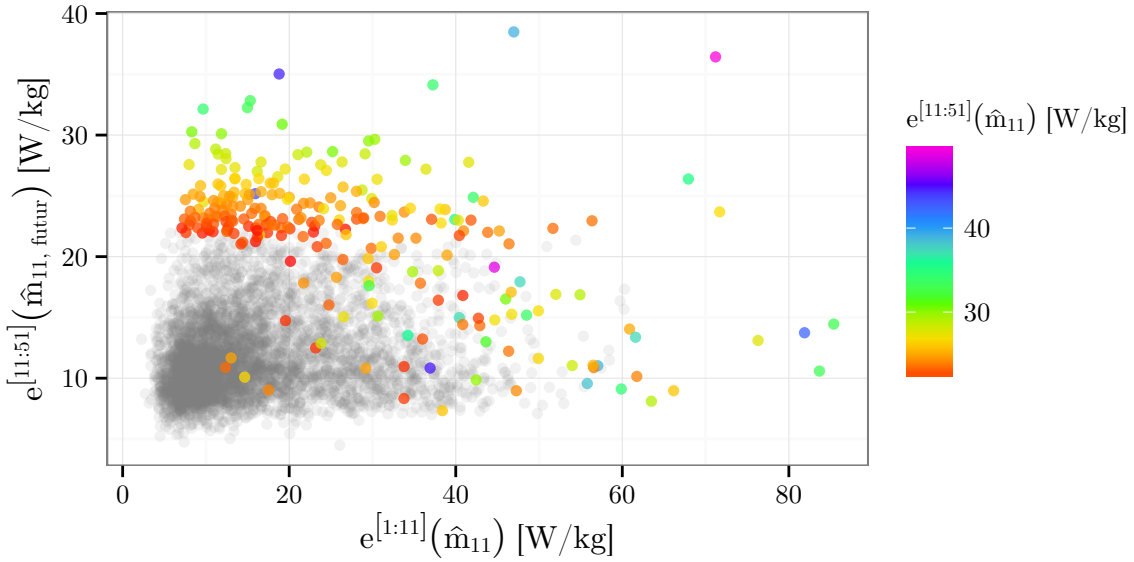


FIGURE 5.22 – Cette figure présente l'erreur $e^{[11:51]}(\hat{m}_{11})$ en fonction de la qualité d'ajustement sur les points passés $e^{[1:11]}(\hat{m}_{11})$ et les points futurs $e^{[11:51]}(\hat{m}_{11,futur})$. Le jeu de trajectoire Mode-C est utilisé. Les $e^{[11:51]}(\hat{m}_{11})$ n'étant pas dans les 95% les plus élevées sont grisées. Seules les 5% les plus élevés sont colorés.

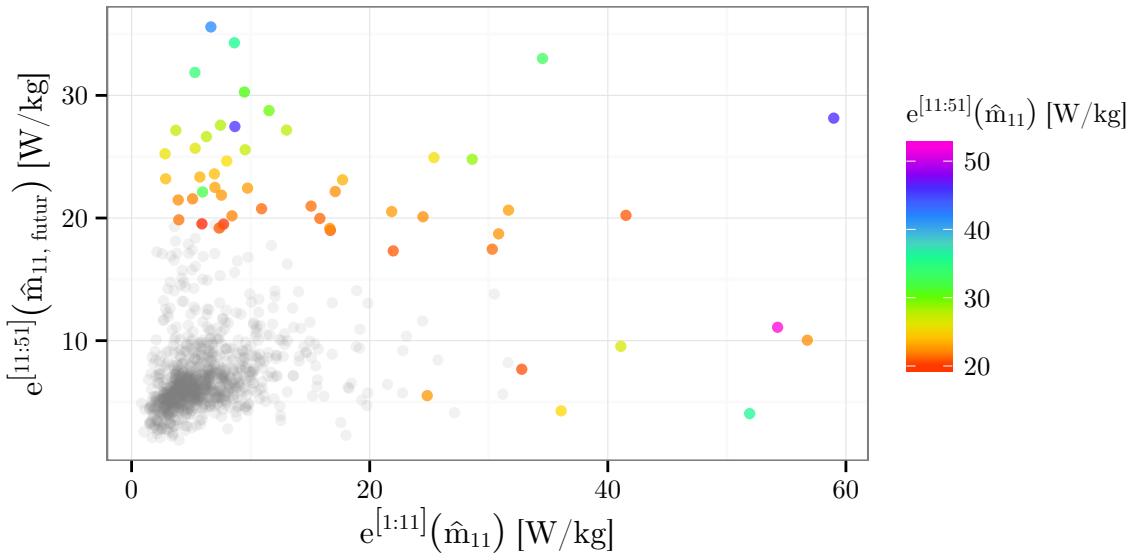


FIGURE 5.23 – Cette figure présente l'erreur $e^{[11:51]}(\hat{m}_{11})$ en fonction de la qualité d'ajustement sur les points passés $e^{[1:11]}(\hat{m}_{11})$ et les points futurs $e^{[11:51]}(\hat{m}_{11,futur})$. Le jeu de trajectoire Mode-S est utilisé. Les $e^{[11:51]}(\hat{m}_{11})$ n'étant pas dans les 95% les plus élevées sont grisées. Seules les 5% les plus élevés sont colorés.

ces méthodes d'estimation de la masse sur les 11 points passés de nos trajectoires, on réduit la RMSE sur l'altitude à $t = 600$ s de 40 % à 50 %. Pour que les erreurs sur la prédiction ne soient dues qu'aux erreurs sur la masse, les prédictions sont calculées en utilisant le profil de vitesse air observé sur les points futurs. Il faut toutefois se rappeler que dans le contexte de la prédiction de trajectoire, ce profil de vitesse air n'est normalement pas connu.

Les masses estimées sont distribuées différemment entre les deux jeux de trajectoires. Elles sont moins dispersées sur les trajectoires Mode-S que les trajectoires Mode-C. Le pic de la distribution des masses estimées sur les trajectoires Mode-S est décalé de 2 000 kg par rapport à celui des trajectoires Mode-C. Les altitudes finales sont en moyenne sous-estimées ce qui peut s'expliquer par une sur-estimation de la masse. Ceci se vérifie sur les masses estimées sur les points futurs. Pour les trajectoires Mode-S, certaines masses estimées sur les points futurs sont bien supérieures à la masse maximale de l'A320. Ceci pourrait s'expliquer par une commande de poussée inférieure à *max climb*.

La qualité de l'ajustement fait lors de l'estimation des masses sur les points passés et sur les points futurs permet d'analyser la cause des erreurs faites sur la prédiction. Un mauvais ajustement sur les points futurs implique que la puissance calculée sera très différente de celle observée. Ceci rend possible de grands écarts au niveau de la trajectoire. Ce mauvais ajustement sur les points futurs est probablement dû à une mauvaise modélisation de la puissance découlant d'une mauvaise hypothèse sur la commande de poussée. L'ajustement sur les points futurs minore l'erreur que l'on fera sur les points futurs avec la masse estimée sur les points passés. Pour certaines trajectoires, l'ajustement sur les points futurs est bon mais la masse estimée sur les points passés est loin de celle estimée sur les points futurs. Sur nos trajectoires, plus l'ajustement sur les points passés est mauvais et plus la qualité de la prédiction de la puissance est probablement mauvaise.

Chapitre 6

Utilisation de méthodes d'apprentissage

Sommaire

6.1 Paramètres des méthodes d'apprentissage 126
6.1.1 Méthodes d'apprentissage 127
6.1.2 Variables explicatives 128
6.2 Apprendre de la masse 129
6.2.1 Mise en œuvre 131
6.2.2 Résultats 131
6.2.3 Conclusion 134
6.3 Apprendre une commande de poussée 138
6.3.1 Lien entre la masse et la poussée 138
6.3.2 Évaluation de la qualité d'une commande de poussée 140
6.3.3 Formulation du problème d'optimisation 141
6.3.4 Résolution du problème d'optimisation 141
6.3.5 Résultats 142
6.4 Apprendre le profil de vitesse $(cas, Mach)$ 149
6.4.1 Ajustement du couple $(cas, Mach)$ 149
6.4.2 Apprentissage direct des cas et $Mach$ ajustés 153
6.5 Comparaison des différentes méthodes 161
6.5.1 Apprendre directement l'altitude 161
6.5.2 Prédiction avec le modèle physique BADA 165
6.5.3 Discussion sur les biais possibles 167
6.6 Conclusion 168

Dans le chapitre 5, on utilise un modèle physique pour construire des méthodes permettant d'estimer la masse d'un avion à partir des points passés d'une trajectoire. Les variables intervenant dans ces estimations sont uniquement des variables apparaissant dans le modèle physique. Dans ce chapitre, on utilise des méthodes d'apprentissage en se servant de

jeux de trajectoires et de toutes les variables disponibles. L'utilisation de ces méthodes est décrite dans la section 6.1. Ces méthodes sont appliquées pour apprendre une masse (section 6.2), une commande de poussée (section 6.3) et un profil de vitesse (section 6.4). Tous ces paramètres sont nécessaires à l'utilisation du modèle physique BADA. Cette approche est illustrée par la figure 6.1. Les résultats obtenus avec tous ces paramètres appris sont décrits dans la section 6.5. On compare ces résultats à ceux obtenus en apprenant directement l'altitude après 10 min de montée ainsi qu'aux résultats obtenus par la modèle BADA utilisé avec les paramètres de référence ou les masses estimées dans le chapitre 5.

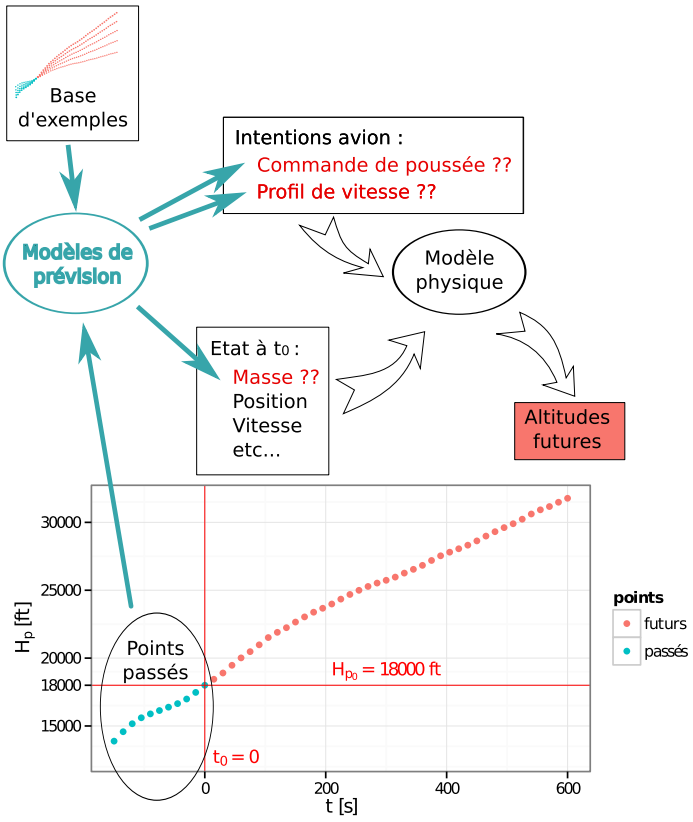


FIGURE 6.1 – Cette figure illustre l'approche consistant à apprendre chacun des paramètres manquants et nécessaires à l'utilisation de BADA.

6.1 Paramètres des méthodes d'apprentissage

Cette section présente les méthodes d'apprentissage et variables explicatives utilisées dans cette thèse.

6.1.1 Méthodes d'apprentissage

Les différentes méthodes testées dans cette thèse sont présentées dans le tableau 6.1. Outre la régression ridge (Ridge), les réseau de neurones (NNet) et les Gradient Boosting Machine (GBM) présentées dans les sections 3.2.1, 3.2.2 et 3.2.3 respectivement ; on a également testé la régression linéaire avec une sélection de variables préalable.

La régression sur composantes principales (PCR) applique une analyse en composantes principales sur les variables explicatives puis ces composantes sont ordonnées selon leur variance. Le modèle prédictif final est construit en appliquant une régression linéaire avec les k composantes de plus grande variance, k étant un hyper-paramètre de la méthode.

On a également testé une *forward selection* mise en œuvre avec une régression linéaire (MLR-FS). Ceci permet d'ordonner les variables explicatives. De même que pour la PCR, le modèle prédictif final est construit en appliquant une régression linéaire avec les k premières variables, k étant un hyper-paramètre de la méthode.

Le réseau de neurones utilisé est un réseau de neurones à une couche cachée. On utilise l'implémentation faite dans le package `nnet` de l'environnement `R`. Pour la méthode GBM, on utilise le package `gbm` de l'environnement `R`.

TABLE 6.1 – Ce tableau résume les différents algorithmes d'apprentissage utilisés et la grille d'hyper-paramètres associée si nécessaire.

méthode	grille d'hyper-paramètres
MLR-FS_k	$k = [\![2; min(120, n_{var})]\!]$
Ridge_λ	$\lambda = 10^{[\![-5;1]\!]} \cup 0.5 \times 10^{[\![-5;0]\!]}$
PCR_k	$k = [\![2; min(120, n_{var})]\!]$
$\text{NNet}_{(n,\lambda)}$	$n = \{2, 3, 4, 5, 6\}$ $\lambda = \{0.1, 0.2, 0.3, 0.4, 0.5\}$
$\text{GBM}_{(m,J,\nu)}$	$m = \{1000, 1500, 2000\}$ $J = \{3, 5, 10, 15\}$ $\nu = \{0.001, 0.0025, 0.005, 0.01, 0.025, 0.05\}$

Pour toutes ces méthodes, il faut sélectionner un jeu d'hyper-paramètres. Cette sélection est faite en se servant uniquement de l'*ensemble d'apprentissage* T. Le choix de l'hyper-paramètre λ est fait en utilisant une *grid search* qui recherche le λ parmi un ensemble de λ définie *a priori*. Le λ choisi est celui minimisant l'erreur sur un sous-ensemble d'exemples issus de T. Cette *grid search* est une implémentation simple du principe décrit dans la section 3.1.3. Dans l'algorithme 6.1, les fonctions $TuneGrid$ et $TuneGridCV$ sont décrites. Ces fonctions embarquent une méthode de sélection d'hyper-paramètre et renvoie le modèle final obtenu à partir de T. $TuneGrid$ sélectionne le λ minimisant l'erreur en *validation simple* et $TuneGridCV$ minimise l'erreur en *validation croisée*. Ainsi $TuneGridCV$ est plus coûteux en nombre d'exécutions de $\mathcal{A}_\lambda$ mais peut s'avérer intéressant si l'on a peu d'exemples et si l'algorithme $\mathcal{A}_\lambda$ s'exécute rapidement. Pour ces raisons $TuneGridCV$ est préféré à $TuneGrid$ pour appliquer les algorithmes MLR-FS_k, Ridge_λ et PCR_k aux données Mode-S.

1 : **fonction** TuneGrid($\mathcal{A}_\lambda$,$grid$)$[T]$
2 : $(T_T, T_V) \leftarrow partition_{(80\%,20\%)}(T)$
3 : $\lambda^* \leftarrow \underset{\lambda \in grid}{argmin}\ Err_{val}(\mathcal{A}_\lambda, T_T, T_V)$
4 : **renvoyer** $\mathcal{A}_{\lambda^*}[T]$
5 : **fin fonction**

1 : **fonction** TuneGridCV($\mathcal{A}_\lambda$,$grid$)$[T]$
2 : $\lambda^* \leftarrow \underset{\lambda \in grid}{argmin}\ CV_{10}(\mathcal{A}_\lambda, T)$
3 : **renvoyer** $\mathcal{A}_{\lambda^*}[T]$
4 : **fin fonction**

Algorithme 6.1 : Les algorithmes *TuneGrid* et *TuneGridCV* choisissent les meilleurs hyper-paramètres au regard du jeu d'exemples en entré et de leurs critères de sélection. Ces algorithmes sont des algorithmes d'apprentissage sans hyper-paramètre obtenus à partir d'une grille d'hyper-paramètres *grid* et d'un algorithme d'apprentissage avec hyper-paramètres $\mathcal{A}_\lambda$. Pour *TuneGrid*, le critère est l'erreur en validation Err_{val} sur 20 % des exemples qui ont été mis de côté. Pour *TuneGridCV*, le critère est l'erreur en obtenu par validation-croisée sur 10 plis.

6.1.2 Variables explicatives

Les variables explicatives sont décrites dans le tableau 6.2. On considère plusieurs jeux de variables explicatives décrits dans le tableau 6.3. Ces différents jeux permettent d'évaluer les performances obtenues suivant les variables disponibles dans un contexte opérationnel. On compte six jeux de variables distincts : r correspond aux variables issues de mesures radar, p les variables réelles issues du plan de vol, c les variables catégorielles du plan de vol, $\hat{m}$ les masses estimées sur les points passés, w les variables issues de la grille météorologique qui est uniquement disponible pour les données Mode-C et t la variante de l'appareil qui est uniquement disponible pour les données Mode-S. La concaténation de ces jeux de variables permet de former d'autres jeux de variables.

TABLE 6.2 – Ce tableau décrit les variables utilisées par les méthodes d'apprentissage.

variables	description
H_p	altitude géopotentielle
V_g	vitesse sol
V_a	vitesse air
V_{aXY}	vitesse air dans le plan (X,Y)
ΔT	différentiel de température (cf. [Nuic 11])
W	wind
W_{along}	vent selon V_{aXY}
W_{across}	vent perpendiculaire à V_{aXY}
θ_c	angle de dérive
CAS	Calibrated Air Speed
$Mach$	nombre de Mach
$1/r_{sol}$	courbure par rapport au sol
$1/r_{air}$	courbure par rapport à l'air
ϕ	angle de roulis
$e = V_a \frac{dV_a}{dt} + g_0 \frac{T}{T-\Delta T} \frac{dH_p}{dt}$	variation d'énergie spécifique
$ew = e + \overrightarrow{W}.\overrightarrow{V_a}$	variation d'énergie spécifique corrigée de l'effet du vent
$\Delta T\,(weather\ grid)$	différentiel de température sur la grille pour différent H_p
$W_{along}\,(weather\ grid)$	vent selon TAS_{XY} sur la grille pour différent H_p
RFL	niveau de vol demandé dans le plan de vol
Speed	vitesse demandée dans le plan de vol
distance	distance entre les deux aéroports
AO	compagnie aérienne
DEP	aéroport de départ
ARR	aéroport d'arrivée
$\hat{m}_{LS}$	masse estimée par méthode des moindres carrés sur les points passés
$\hat{m}_{AD}$	masse estimée par méthode adaptative sur les points passés
t	variante de l'appareil

6.2 Apprendre de la masse

Dans la section 5.6 on a étudié des méthodes d'estimation de la masse reposant sur le modèle physique BADA. En supposant une poussée *max climb*, la masse $\hat{m}_{11}$ est estimée en utilisant les points passés de la trajectoire considérée. Sur nos jeux de trajectoires, les méthodes d'estimation de masses décrites précédemment surestiment la masse. Cet écart entre ces deux estimations suggèrent une erreur dans la modélisation des forces et/ou dans l'hypothèse faite sur la commande de poussée. Tant sur les points passés que sur les points futurs, les variations sur la variation d'énergie spécifique observée montrent clairement pour

TABLE 6.3 – Ce tableau résume les différents jeux de variables utilisés.

variables	nom du groupe	nature	source
H_p $\frac{dH_p}{dt}$ $\frac{d^2H_p}{dt^2}$ V_g $\frac{dV_g}{dt}$ V_a $\frac{dV_a}{dt}$ V_{aXY} ΔT W W_{along} W_{across} $\frac{dW_{along}}{dt}$ θ_c CAS $\frac{dCAS}{dt}$ $Mach$ $\frac{dMach}{dt}$ $1/r_{sol}$ $1/r_{air}$ ϕ e ew	r	réelle	mesure
RFL Speed distance	p	réelle	plan de vol
AO DEP ARR	c	catégorielle	plan de vol
$\Delta T\,(weather\ grid)$ $W_{along}\,(weather\ grid)$	w	réelle	Météo-France
$\hat{m}_{LS}$ $\hat{m}_{AD}$	$\hat{m}$	réelle	estimation
t	t	catégorielle	adresse Mode-S 24 bits

certaines trajectoires que l'hypothèse *max climb* est fausse. Malgré ces erreurs, la masse $\hat{m}_{11,futur}$ permet de générer des trajectoires proches de celles observées, comme le montrent les tableaux 5.4 et 5.5. Malheureusement, la masse $\hat{m}_{11,futur}$ n'est pas connue au moment du calcul de la prédiction, elle est calculée *a posteriori* sur la trajectoire future observée.

On peut toutefois construire un jeu d'exemples contenant la masse $\hat{m}_{11,futur}$ à partir d'un jeu de trajectoires dont on connaît les points passés et points futurs. Dans cette sous-section, on exploite le jeu d'exemples ainsi construit avec des méthodes d'apprentissage artificiel. Ces méthodes permettent d'extraire une fonction reliant les paramètres connus dans le passé à la masse $\hat{m}_{11,futur}$ ajustée sur la trajectoire future. On sait que cette masse fournit une bonne prédiction avec le modèle BADA et l'hypothèse *max climb*. Les méthodes d'apprentissage permettent de s'abstraire d'une modélisation physique des forces et de prendre en compte des paramètres qui n'apparaissent normalement pas dans la modélisation des forces comme par exemple la distance à parcourir ou la compagnie aérienne. Ces paramètres sont directement liés à la masse au décollage et à l'exploitation de l'appareil.

Dans cette section, différentes méthodes d'apprentissage ainsi que différents jeux de variables explicatives sont testés sur les trajectoires Mode-C et Mode-S. La qualité du modèle prédictif appris sera évaluée au regard de l'écart entre la masse prédite $\tilde{m}_{11,futur}$ et la masse $\hat{m}_{11,futur}$. Comme la prédiction de cette masse se fait dans le but de prédire les trajectoires, on évalue aussi la qualité de la trajectoire prédite obtenue avec la masse $\tilde{m}_{11,futur}$ en prenant l'hypothèse *max climb* et en utilisant le profil de vitesse observé sur les points futurs. On se place ainsi dans le même cadre que la section 5.6.

6.2.1 Mise en œuvre

Le tableau 6.1 de la section 6.1 présente tous les algorithmes d'apprentissage testés et la grille d'hyper-paramètres associée suivant l'algorithme. Les algorithmes $TuneGrid$ ou $TuneGridCV$ décrits dans 6.1 permettent de choisir les hyper-paramètres des algorithmes d'apprentissage. Les statistiques présentées sont calculées en se servant des prédictions obtenues par $CV_{10}(TuneGrid\left(\mathcal{A}_\lambda, grid\right), S)$ ou $CV_{10}(TuneGridCV\left(\mathcal{A}_\lambda, grid\right), S)$.

6.2.2 Résultats

Les algorithmes d'apprentissage permettent d'apprendre une fonction donnant la masse future estimée $\hat{m}_{11,futur}$ en fonction des variables connues au moment où la prédiction de trajectoire est calculée. Les tableaux 6.4 et 6.5 présentent les statistiques associées à la prédiction de $\hat{m}_{11,futur}$. Sur ces tableaux, les modèles linéaires ont des performances similaires. De même, les NNet (réseau de neurones) et les GBM (Gradient Boosting Machine) ont des performances comparables avec un léger avantage pour les GBM. Pour toutes les méthodes, les erreurs se réduisent en ajoutant des variables. Toutefois, l'erreur n'est pas significativement réduite par l'ajout des variables météo w. La réduction la plus importante est obtenue par l'ajout des masses estimées $\hat{m}$. Ceci est attendu car ces variables sont hautement corrélées à $\hat{m}_{11,futur}$ avec des coefficients de corrélation supérieurs à 0,94. En comparaison, toutes les variables de r ont une corrélation avec $\hat{m}_{11,futur}$ inférieure à 0,61.

Toutefois, ces coefficients de corrélations sont calculés en prenant les variables séparément. Or, plusieurs variables inutiles toutes seules peuvent se révéler utiles ensembles dans le contexte de la régression ([Guyon 03]).

TABLE 6.4 – Statistiques, en kilogrammes, sur l'écart entre la masse apprise et la masse ajustée sur les points futurs, pour les trajectoires Mode-C. Ces statistiques sont calculées sur l'ensemble des valeurs $\tilde{m}_{11,futur} - \hat{m}_{11,futur}$. Les premières lignes reprennent, pour comparaison, les masses estimées du tableau 5.6 de la section 5.6.

variables	méthode	moy	écart-type	moy abs	rmse	max abs
-	$mass_{ref}$	-119	4703	3751	4704	15397
-	naïve	1524	2050	1868	2554	27281
-	adaptative	747	2054	1712	2186	15241
-	moindres carrés	1552	2034	1868	2558	23335
-	robuste	1452	1825	1773	2332	24505
r	MLR-FS	3.63	1675	1200	1675	24307
r	Ridge	6.91	1684	1204	1684	24519
r	PCR	5.29	1777	1302	1777	25155
r	NNet	3.45	1466	1069	1466	18474
r	GBM	9.95	1455	1081	1454	13799
$\hat{m}r$	MLR-FS	4.29	1542	1115	1542	20610
$\hat{m}r$	Ridge	4.83	1545	1118	1545	21290
$\hat{m}r$	PCR	5.78	1547	1117	1547	21013
$\hat{m}r$	NNet	2.25	1421	1053	1421	13574
$\hat{m}r$	GBM	2.54	1412	1052	1412	14438
$\hat{m}rw$	MLR-FS	10	1537	1114	1536	20757
$\hat{m}rw$	Ridge	4.36	1539	1114	1539	21439
$\hat{m}rw$	PCR	5.18	1539	1113	1538	21279
$\hat{m}rw$	NNet	-1.81	1406	1041	1405	14045
$\hat{m}rw$	GBM	10.1	1407	1047	1407	14362
$\hat{m}prw$	MLR-FS	7.15	1498	1085	1498	20134
$\hat{m}prw$	Ridge	4.72	1504	1086	1504	21106
$\hat{m}prw$	PCR	5.61	1508	1088	1508	21105
$\hat{m}prw$	NNet	-2.63	1377	1013	1377	14781
$\hat{m}prw$	GBM	5.54	1350	996	1350	14313
$c\hat{m}prw$	GBM	6.78	1303	953	1303	14730

Les tableaux 6.6 et 6.7 présentent les résultats obtenus concernant la prédiction de trajectoire. Les remarques faites précédemment concernant la masse prédite sont aussi valables pour l'altitude prédite. La plus faible RMSE est obtenue en utilisant GBM avec le jeu de variables le plus complet. En comparaison de la prédiction obtenue avec la masse de référence m_{ref}, la $RMSE$ est réduite de 70 %. Elle est réduite de 50 % comparée à celle obtenue avec la méthode adaptative. La masse prédite permet de réduire la part de

TABLE 6.5 – Statistiques, en kilogrammes, sur l'écart entre la masse apprise et la masse ajustée sur les points futurs, pour les trajectoires Mode-S. Ces statistiques sont calculées sur l'ensemble des valeurs $\tilde{m}_{11,futur} - \hat{m}_{11,futur}$. Les premières lignes reprennent, pour comparaison, les masses estimées du tableau 5.7 de la section 5.6.

variables	méthode	moy	écart-type	moy abs	rmse	max abs
-	$mass_{ref}$	1993	3412	3079	3950	14825
-	naïve	1436	1868	1752	2356	22343
-	adaptative	1274	1634	1673	2071	16205
-	moindres carrés	1395	1887	1716	2346	24791
-	robuste	1348	1680	1657	2153	21517
r	MLR-FS	8.04	1914	1341	1913	16808
r	Ridge	17.6	1885	1325	1884	15265
r	PCR	17	1882	1327	1881	14563
r	NNet	22.7	1695	1176	1694	20768
r	GBM	39.3	1449	1027	1449	15498
$\hat{m}r$	MLR-FS	2.23	1472	1011	1471	16601
$\hat{m}r$	Ridge	-5.35	1560	1065	1559	17701
$\hat{m}r$	PCR	11.7	1564	1060	1563	17813
$\hat{m}r$	NNet	16.3	1553	1084	1553	14522
$\hat{m}r$	GBM	25.7	1432	988	1431	16917
$\hat{m}pr$	MLR-FS	2.23	1472	1011	1471	16601
$\hat{m}pr$	Ridge	-4.89	1561	1069	1561	17632
$\hat{m}pr$	PCR	-1.87	1546	1052	1546	17737
$\hat{m}pr$	NNet	-16.1	1564	1078	1563	16429
$\hat{m}pr$	GBM	31.3	1422	973	1421	16867
$c\hat{m}pr$	GBM	19.6	1413	966	1412	16911
$c\hat{m}prt$	GBM	16.7	1342	891	1342	16385

l'erreur $e^{[11:51]}(m)$ due à une mauvaise masse (voir figure 6.2). L'erreur due à une mauvaise modélisation de l'excès de puissance spécifique demeure incompressible : on ne peut faire mieux que $e^{[11:51]}(\hat{m}_{11,futur})$.

TABLE 6.6 – Statistiques, en pieds, sur l'écart en altitude entre la trajectoire prédite avec la masse apprise et le dernier plot observé pour les trajectoires Mode-C. Ces statistiques sont calculées sur l'ensemble des valeurs $\left(H_p^{(pred)}(\tilde{m}_{11,futur}) - H_p^{(obs)}\right)$ ($t = 600$ s). Les trajectoires sont calculées avec $V_a(t) = V_a{}^{(obs)}(t)$.

variables	méthode	moy	écart-type	moy abs	rmse	max abs
-	$mass_{ref}$	-82.4	1479	1168	1481	5493
-	naïve	-524	651	628	836	6874
-	adaptative	-303	685	582	749	5535
-	moindres carrés	-532	653	631	843	6033
-	robuste	-503	595	603	779	6843
r	MLR-FS	-69.4	564	413	568	6884
r	Ridge	-70.2	566	414	571	6798
r	PCR	-70	589	437	593	6948
r	NNet	-66.7	505	376	510	5294
r	GBM	-71.7	500	377	505	5045
$\hat{m}r$	MLR-FS	-68.6	527	391	532	5842
$\hat{m}r$	Ridge	-68.7	527	391	532	6012
$\hat{m}r$	PCR	-69.3	528	391	532	5943
$\hat{m}r$	NNet	-67.2	495	373	500	4968
$\hat{m}r$	GBM	-68	489	370	494	5263
$\hat{m}rw$	MLR-FS	-70.6	526	391	530	5879
$\hat{m}rw$	Ridge	-68.5	526	390	530	6049
$\hat{m}rw$	PCR	-69.2	525	390	530	6009
$\hat{m}rw$	NNet	-66.2	490	369	494	5129
$\hat{m}rw$	GBM	-70.4	488	369	493	5238
$\hat{m}prw$	MLR-FS	-69.6	514	381	519	5721
$\hat{m}prw$	Ridge	-68.4	515	380	519	5966
$\hat{m}prw$	PCR	-69.2	516	381	520	5966
$\hat{m}prw$	NNet	-65.4	479	360	484	5291
$\hat{m}prw$	GBM	-68.7	470	353	475	5221
$c\hat{m}prw$	GBM	-68.8	452	337	457	5363

6.2.3 Conclusion

Les méthodes d'estimation de la masse se basent sur la modélisation physique des forces faite par BADA. Aucun jeu de trajectoires n'est utilisé. Seuls les points passés de la trajectoire considérée sont exploités. Pour cela, une commande de poussée pour ces points

TABLE 6.7 – Statistiques, en pieds, sur l'écart en altitude entre la trajectoire prédite avec la masse apprise et le dernier plot observé pour les trajectoires Mode-S. Ces statistiques sont calculées sur l'ensemble des valeurs $\left(H_p^{(pred)}(\tilde{m}_{11,futur}) - H_p^{(obs)}\right)$ ($t = 600$ s). Les trajectoires sont calculées avec $V_a(t) = {V_a}^{(obs)}(t)$.

variables	méthode	moy	écart-type	moy abs	rmse	max abs
-	$mass_{ref}$	-810	1133	1079	1392	5531
-	naïve	-598	602	673	849	6574
-	adaptative	-551	554	651	781	4171
-	moindres carrés	-584	606	663	842	7150
-	robuste	-571	550	645	792	6371
r	MLR-FS	-156	633	472	652	5603
r	Ridge	-156	624	466	643	5134
r	PCR	-159	622	468	642	5028
r	NNet	-157	569	438	590	5759
r	GBM	-164	493	392	519	3824
$\hat{m}r$	MLR-FS	-151	500	391	522	4214
$\hat{m}r$	Ridge	-145	529	407	548	4610
$\hat{m}r$	PCR	-154	534	408	555	4651
$\hat{m}r$	NNet	-155	530	408	552	4060
$\hat{m}r$	GBM	-159	489	387	514	4327
$\hat{m}pr$	MLR-FS	-151	500	391	522	4214
$\hat{m}pr$	Ridge	-146	530	409	549	4585
$\hat{m}pr$	PCR	-150	529	406	550	4623
$\hat{m}pr$	NNet	-145	536	412	555	4152
$\hat{m}pr$	GBM	-161	486	385	512	4309
$c\hat{m}pr$	GBM	-157	482	382	507	4325
$c\hat{m}prt$	GBM	-155	451	351	477	4137

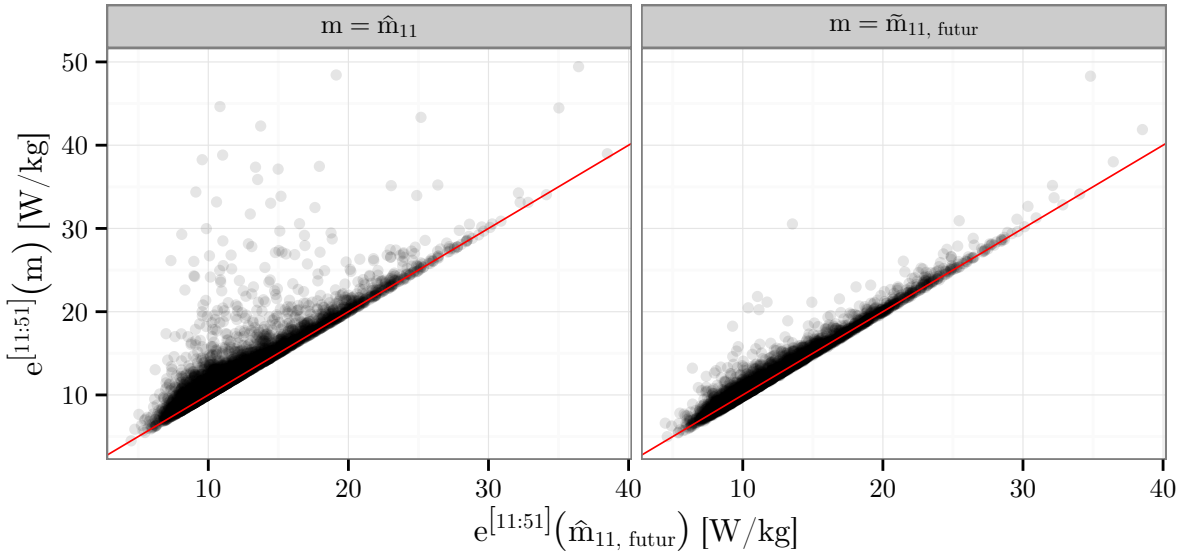


FIGURE 6.2 – Cette figure présente l'erreur $e^{[1:51]}(m)$ obtenue avec la masse m (estimée à gauche, apprise à droite) en fonction de l'erreur minimale $e^{[1:51]}(\hat{m}_{11,futur})$. La masse $\hat{m}_{11}$ est estimée par moindres carrés et la masse $\tilde{m}_{11,futur}$ est celle prédite par GBM. La droite rouge a pour équation $y = x$. Elle permet de lire l'écart entre l'erreur obtenue et l'erreur minimale. Cette figure présente uniquement les trajectoires Mode-C.

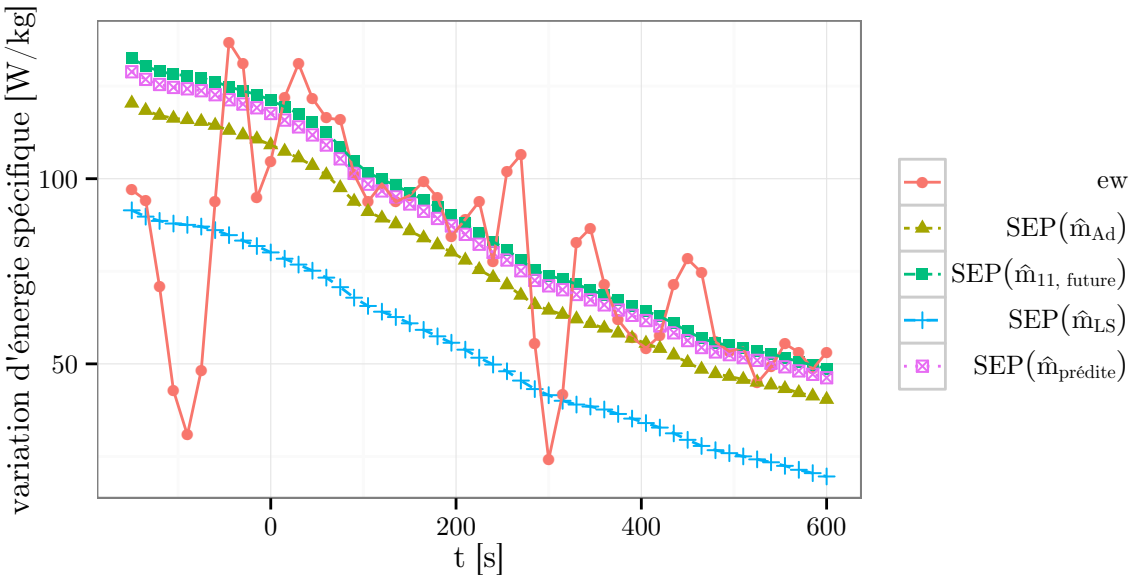


FIGURE 6.3 – Cette figure présente l'excès de puissance spécifique calculée avec différentes masses (SEP) et la variation d'énergie spécifique observée (ew). Un seul avion est considéré. Selon les lois de Newton, l'excès de puissance spécifique est égal à la variation d'énergie spécifique.

passés est supposée *a priori*. Cette supposition peut être fausse comme illustré par la figure 6.3.

L'apprentissage artificiel n'exploite pas un modèle physique et ne nécessite aucune hypothèse *a priori* pour prédire la masse. À la place, elle utilise un jeu d'exemples $(y_i, x_i)_{1 \leqslant i \leqslant n}$ pour construire un modèle qui relie la variable à prédire y aux variables explicatives x. Dans notre application les points « futurs » des exemples de trajectoires sont utilisés pour construire le modèle qui calcule notre y, ici la masse qui donnera la meilleure prédiction d'altitude pour les points futurs. Ainsi, l'algorithme d'apprentissage va exploiter tous les points d'un jeu de trajectoires pour construire le modèle.

Dans les deux approches, la prédiction de la masse est faite à l'aide de variables disponibles au moment du calcul de la trajectoire prédite. Toutefois, le modèle construit à partir d'un algorithme d'apprentissage peut prendre en compte des variables n'apparaissant pas dans un modèle physique, comme la distance entre l'aéroport de départ et l'aéroport d'arrivée. Pourtant la masse au décollage est bien évidemment liée à cette distance à parcourir.

Les méthodes d'apprentissage prédisent mieux la masse que les méthodes d'estimation de la masse. Ceci peut s'expliquer par l'emploi d'un jeu de trajectoires et de variables inutilisées par le modèle physique.

6.3 Apprendre une commande de poussée

Dans la section 5, les méthodes d'estimation de la masse reposent sur l'hypothèse d'une poussée *max climb*. Cette hypothèse s'avère souvent fausse *a posteriori*. Dans cette section, on se propose d'apprendre une commande de poussée d'un jeu de trajectoires observées. Cette commande est construite dans le but de réduire l'écart entre l'excès de puissance spécifique modélisé et la variation d'énergie spécifique qui est finalement observée. Elle est commune à tous les avions. Elle est apprise sur un jeu d'apprentissage et testée sur un jeu de validation. Pour ce faire, on utilise la validation croisée à 10 plis. Cette section reprend les travaux publiés dans [Alligier 13b].

6.3.1 Lien entre la masse et la poussée

Pour prendre en compte une commande de poussée, on modifie le calcul de la puissance en introduisant un coefficient c dans le calcul de la poussée :

$$Puiss\left(c, m\right) = \left(cThr_{max\ climb} - D\left(m\right)\right) V_a \tag{6.1}$$

$$SpecPuiss\left(c, m\right) = \frac{Puiss\left(c, m\right)}{m} \tag{6.2}$$

$$P\left(c, m\right) = Puiss\left(c, m\right) V_a - mQ \tag{6.3}$$

Avec cette nouvelle modélisation, la figure 6.4 permet de visualiser l'excès de puissance spécifique lorsque la poussée n'est pas *max climb*. On remarque que la commande

de poussée a une forte influence sur l'excès de puissance spécifique. On a par exemple $SpecPuiss\,(1.009, 65\ 000\ \text{kg}) = SpecPuiss\,(1, 64\ 000\ \text{kg})$. Ainsi, un écart de moins de 1 % sur la commande de poussée peut compenser une masse supérieure de 1 000 kg. Ceci montre l'importance de l'hypothèse de poussée dans les méthodes d'estimation de la masse.

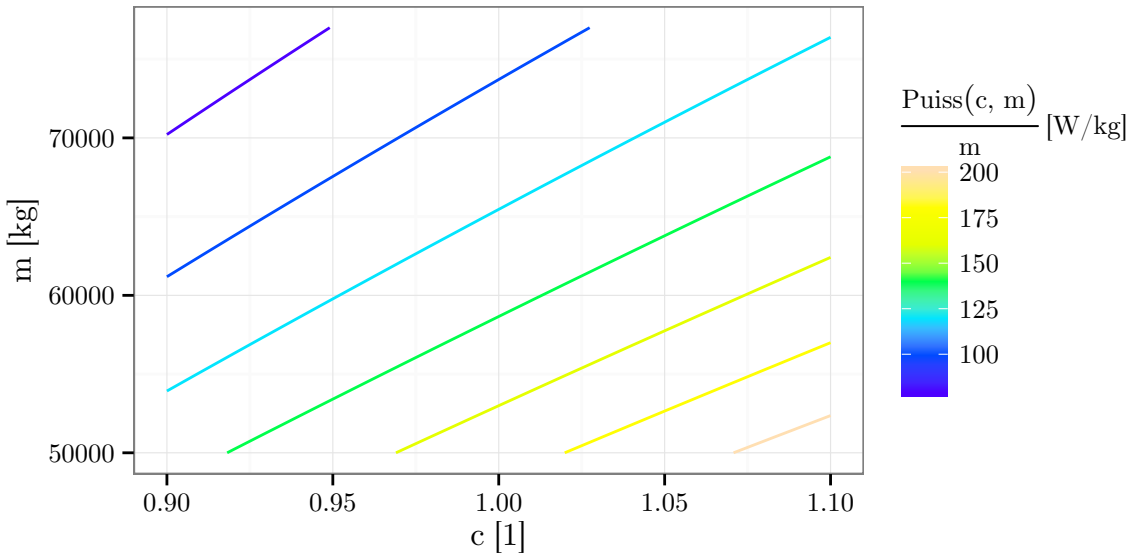


FIGURE 6.4 – Excès de puissance spécifique en fonction de la masse m et de la commande de poussée de c. Le calcul est fait à l'altitude $H_p = 18\ 000$ ft et à la vitesse $cas = 310$ kts dans une atmosphère ISA.

Avec l'égalité précédente, on remarque également que deux couples (c, m) différents peuvent générer le même excès de puissance spécifique. Ceci est vrai concernant le calcul en un point mais également pour un profil d'excès de puissance spécifique complet. Il n'y a pas de modèle contraignant la variation de la commande de poussée. Ainsi, sans contraintes, rien n'empêche pour n'importe quelle séquence de masses $(m_i)_{1 \leqslant i \leqslant n}$ donnée, de choisir la séquence de commande $C = (C_1, \ldots, C_n)$ telle que $P_i\,(C_i, m_i) = m_i Q_i$ i.e. telle que l'excès de puissance spécifique calculé égalise la variation d'énergie spécifique observée. La figure 6.5 donne un exemple pour une trajectoire donnée et deux séquences de masses distinctes.

Il y a une infinité de couples (C, m) expliquant parfaitement les observations. Ainsi, lorsque l'on trouve un couple (C, m) expliquant parfaitement la variation d'énergie spécifique observée, on ne peut raisonnablement pas conclure que les paramètres C et m trouvés sont proches des paramètres réels de l'avion. Lorsque l'on ne s'intéresse qu'à la trajectoire de l'avion, ceci n'est pas gênant car chacun de ces couples (C, m) permet d'expliquer parfaitement la trajectoire. De ce point de vue, tous ces couples se valent.

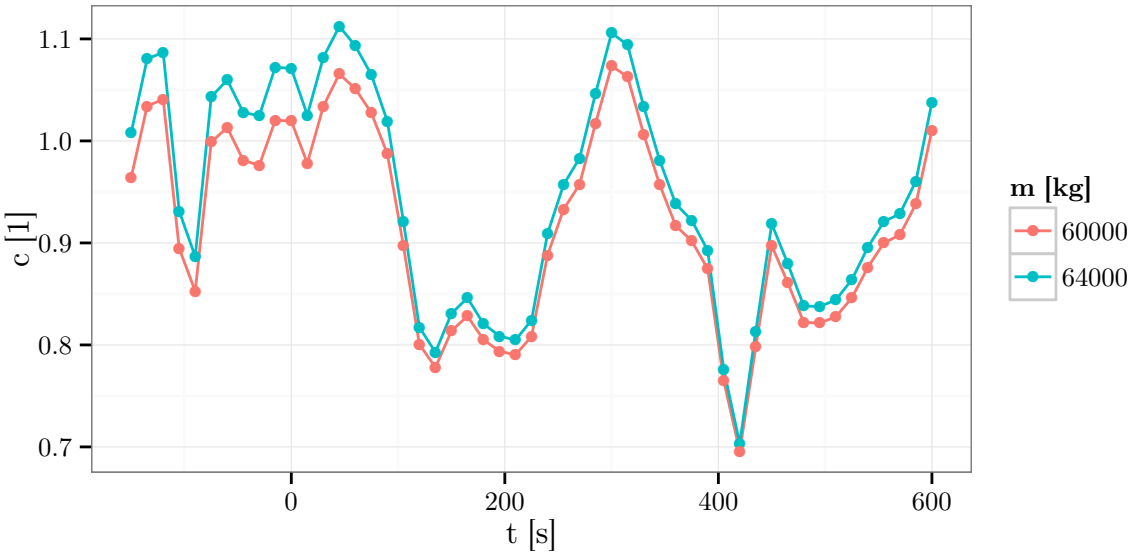


FIGURE 6.5 – Ces deux commandes de poussée associée chacune à une masse différente expliquent parfaitement la variation d'énergie spécifique observée d'une trajectoire Mode-C donnée.

6.3.2 Évaluation de la qualité d'une commande de poussée

Chaque avion est associé a une famille de couples (C, m) équivalents. La notion de distance entre un profil de poussée donné et une famille de profils (C, m) associé à un avion k est délicate à appréhender. Comme l'on cherche à diminuer l'erreur sur l'excès de puissance spécifique, on peut considérer l'erreur $\mathcal{E}_k$ faite sur n points pour l'avion k de masse m_k au dernier point :

$$\mathcal{E}_k\left(\left(C_1, \ldots, C_n\right), m_k\right) = \sum_{i=1}^{n} \left(\frac{P_{k,i}\left(C_i, m_k + \delta_{k,i}\right)}{m_k + \delta_{k,avg}} \right)^2 \tag{6.4}$$

Dans cette formule, on reprend les notations précédemment introduite dans la section 5 : $\delta_{k,i}$ est le carburant consommé entre le point i et le point n, $\delta_{k,avg}$ est la moyenne des δ_i sur les n points considérés et m_k est la masse de l'avion au point n. La masse m_k est toutefois inconnue. Si l'on suppose que l'avion suit un profil C, il est logique de se servir de ce profil pour estimer la masse de l'avion : $m_k^*(C) = \underset{m \in]-\delta_{avg}; +\infty[}{argmin} \mathcal{E}_k(C, m)$. Cette masse m_k^* dépend bien sûr du profil C choisi. L'annexe A démontre que la fonction m_k^* existe et est de classe C^∞. Au final, la valeur de $\mathcal{E}_k(C, m_k^*)$ permet d'évaluer l'erreur commise dans le meilleur des cas une fois que l'on a choisi le profil C. Il est à noter que les profils C équivalents auront une erreur $\mathcal{E}_k(C, m_k^*)$ identique.

6.3.3 Formulation du problème d'optimisation

Dans la sous-section précédente, on a construit une grandeur permettant d'évaluer l'adéquation d'un profil à une trajectoire donnée. Si l'on veut construire un profil unique pour toutes les trajectoires on peut se servir du critère précédemment défini pour minimiser l'erreur faite en utilisant ce profil unique sur K trajectoires :

$$\mathcal{E}_{total}\left(C\right) = \sum_{k=1}^{K} \mathcal{E}_k\left(C, m_k^*\right). \tag{6.5}$$

Le profil unique ainsi modélisé impose que les points ayant même indice aient une même commande. Les points sont indicés par rapport à la date à laquelle l'avion franchit $H_p = 18\ 000$ ft. Toutefois, la commande de poussée peut être dictée par d'autres variables. On peut par exemple considérer que les avions ont une même commande de poussée à altitude égale. On peut aussi penser que plusieurs variables interviennent dans le choix de la commande de poussée. Pour prendre en compte ceci, on considère la commande $c = c(x_{k,i}; \theta)$ avec c une fonction paramétrée par le vecteur θ et $x_{k,i}$ le vecteur de variables utiles associé à l'avion k au point i. L'erreur totale $\mathcal{E}_{total}$ se réécrit :

$$\begin{aligned} \mathcal{E}_{total}\left(\theta\right) =& \sum_{k=1}^{K} \mathcal{E}_k\left(C_k\left(\theta\right), m_k^*\right), \\ & \text{avec } C_k\left(\theta\right) = \left(c(x_{k,1}; \theta), \ldots, c(x_{k,n}; \theta)\right). \end{aligned} \tag{6.6}$$

6.3.4 Résolution du problème d'optimisation

Une fois la fonction paramétrique $c(x; \theta)$ choisie, il nous faut trouver les paramètres θ minimisant $\mathcal{E}_{total}$. Pour chaque évaluation de $\mathcal{E}_{total}\left(\theta\right)$, on doit résoudre pour chaque avion un sous-problème de minimisation pour déterminer la masse m_k^*. Ce sous-problème a été résolu de manière efficace dans la sous-section 5.6. Il se ramène à trouver les racines positives d'un polynôme de degré 4. L'évaluation de $\mathcal{E}_{total}$ est ainsi relativement rapide. De plus, le gradient de $\mathcal{E}_{total}$ peut être aisément calculé en déterminant la dérivée partielle de l'erreur $\mathcal{E}_k$ par rapport à la composante j du vecteur θ. En se rappelant que selon l'annexe A m_k^* satisfait l'équation $\frac{\partial \mathcal{E}_k}{\partial m}(C_k, m_k^*) = 0$, la dérivée s'écrit :

$$\begin{aligned} \frac{\partial \mathcal{E}_k}{\partial \theta_j}(\theta) &= \frac{\partial \mathcal{E}_k}{\partial m}(C_k, m_k^*)\ \frac{\partial m_k^*}{\partial \theta_j} + \sum_{i=1}^{n} \frac{\partial \mathcal{E}_k}{\partial C_{k,i}}(C_k, m_k^*)\ \frac{\partial C_{k,i}}{\partial \theta_j} \\ &= \sum_{i=1}^{n} \frac{\partial \mathcal{E}_k}{\partial C_{k,i}}(C_k, m_k^*)\ \frac{\partial c}{\partial \theta_j}\left(x_{k,i}; \theta\right) \\ &= 2 \sum_{i=1}^{n} \frac{P_{k,i}\left(c(x_{k,i}; \theta), m_k^* + \delta_{k,i}\right)}{m_k^* + \delta_{k,avg}} \frac{V_{ak,i}\, Thr_{k,i}}{m_k^* + \delta_{k,avg}} \frac{\partial c}{\partial \theta_j}\left(x_{k,i}; \theta\right). \end{aligned} \tag{6.7}$$

Ceci permet de conclure sur la dérivée partielle de $\mathcal{E}_{total}$:

$$\frac{\partial \mathcal{E}_{\text{total}}}{\partial \theta_j}(\theta) = \sum_{k=1}^{K} \frac{\partial \mathcal{E}_k}{\partial \theta_j}(\theta). \tag{6.8}$$

Ce gradient est utilisé dans BFGS ([Broyden 70, Fletcher 70, Goldfarb 70, Shanno 70]), une méthode d'optimisation locale utilisant le gradient, pour trouver θ minimisant $\mathcal{E}_{total}(\theta)$.

6.3.5 Résultats

Choix de la fonction paramétrique c

La méthode permettant d'obtenir une commande de poussée est testée sur les jeux de données Mode-C et Mode-S. Avant de mettre en œuvre cette méthode, il faut modéliser la fonction c. Les fonctions (6.9) et (6.10) modélisent des commandes avec un simple polynôme d'une variable et de degré 4.

$$c_{H_p}(x;\theta) = \theta_0 + \theta_1 H_p + \theta_2 {H_p}^2 + \theta_3 {H_p}^3 + \theta_4 {H_p}^4 \tag{6.9}$$

$$c_t(x;\theta) = \theta_0 + \theta_1 t + \theta_2 t^2 + \theta_3 t^3 + \theta_4 t^4 \tag{6.10}$$

Les fonctions testées ici sont simples et on pourrait tester des fonctions plus compliquées comme un réseaux de neurones par exemple. Les variables catégorielles peuvent aussi être utiles. La fonction $c_{variante}$, définie par l'équation (6.11), prend en argument une fonction c paramétrée par un vecteur de longueur 5. Cette fonction $c_{variante}$ permet de construire une commande par variante différente d'A320. Ici, les différentes variantes ont été séparées en trois groupes suivant la motorisation [1]. Les variantes les plus présentes sont les 211, 214 et 232. Le compte des autres variantes ne dépassent pas la dizaine de trajectoires. Chaque groupe contient au moins 170 trajectoires. Chacun de ces groupes de trajectoires sert à apprendre une commande de poussée.

$$\begin{aligned} c_{variante}[c](x;\theta) =& c(x;\theta_{[\![0:4]\!]}) \\ &+ \delta_{variante,\{211,212\}}(x) c(x;\theta_{[\![5:9]\!]}) \\ &+ \delta_{variante,\{214,216\}}(x) c(x;\theta_{[\![10:14]\!]}) \\ &\text{avec } \delta_{variante,num}(x) = \begin{cases} 1 & \text{si } variante(x) \in num \\ 0 & \text{sinon} \end{cases} \end{aligned} \tag{6.11}$$

La conception d'une commande de poussée peut être vu selon deux axes : *1*) le choix de la fonction paramétrique qui présuppose des relations entre les variables explicatives et la variable à expliquer et *2*) le partitionnement des exemples qui permet de choisir ce que représente la variable à expliquer. Ceci permet de choisir le concept que l'on souhaite apprendre. Ainsi, $c_{variante}[c]$ permet d'apprendre la commande de poussée de chaque variante. À contrario, c permet d'apprendre la « moyenne » de la commande de poussée des différentes variantes pondérées par leurs apparitions dans les exemples. Ceci peut ne pas être gênant si la variante n'a pas ou peu d'incidence sur la commande de poussée. De plus,

1. CFM56-5A pour les variantes 211 et 212 ; CFM56-5B pour les variantes 214 et 216 ; IAE Model pour les variantes 131, 231, 232 et 233.

le partitionnement des exemples réduit le nombre d'exemples sur lesquels les commandes sont apprises, ce qui peut dégrader la qualité de l'estimation.

Analyse des commandes apprises

En appliquant notre méthode sur chaque *ensemble d'apprentissage* des 10 plis de la validation croisée, on obtient 10 commandes de poussée.

Les figures 6.6 et 6.7 présentent les commandes de poussée c_t et c_{H_p}. Les variations des commandes c_t et c_{H_p} sont assez similaires pour un même jeu de données. Ceci est probablement dû au fait que l'altitude H_p est croissante avec t. L'amplitude des variations est de 5 % pour les trajectoires Mode-C et de 4 % pour les trajectoire Mode-S. Lorsque l'on compare les commandes obtenues avec les trajectoires Mode-C et les trajectoires Mode-S, on remarque que les commandes n'ont pas du tout la même variation. Ces différences sont délicates à expliquer, elles peuvent par exemple provenir de procédures différentes.

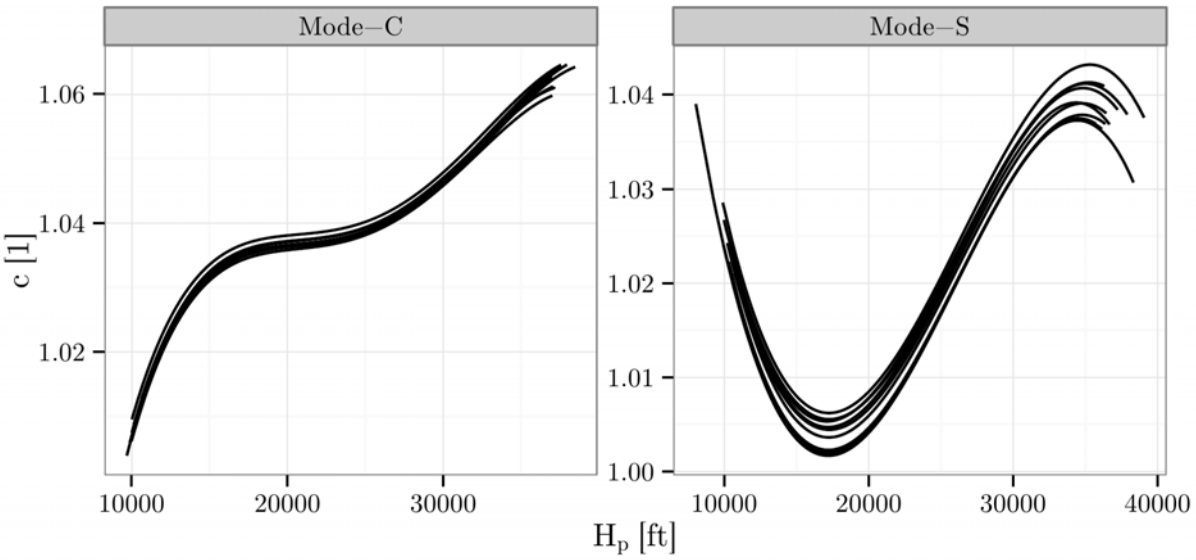


FIGURE 6.6 – Cette figure présente les commandes c_{H_p}. 10 commandes de poussée sont tracées correspondant aux 10 plis de la validation croisée.

Les figures 6.8 et 6.9 présentent les commandes de poussée $c_{variante}[c_t]$ et $c_{variante}[c_{H_p}]$. La variante de l'appareil est disponible uniquement pour les trajectoires Mode-S. On remarque sur ces figures que les commandes ont des variations distinctes suivant la variante de l'appareil. La variante 211 est celle exhibant une commande constante. Pour les deux autres variantes, la variation de la commande de poussée a une amplitude d'environ 5 %. Cette différence est relativement élevée si l'on considère que 1 % de variation de la commande puisse correspondre à une différence de 1 000 kg.

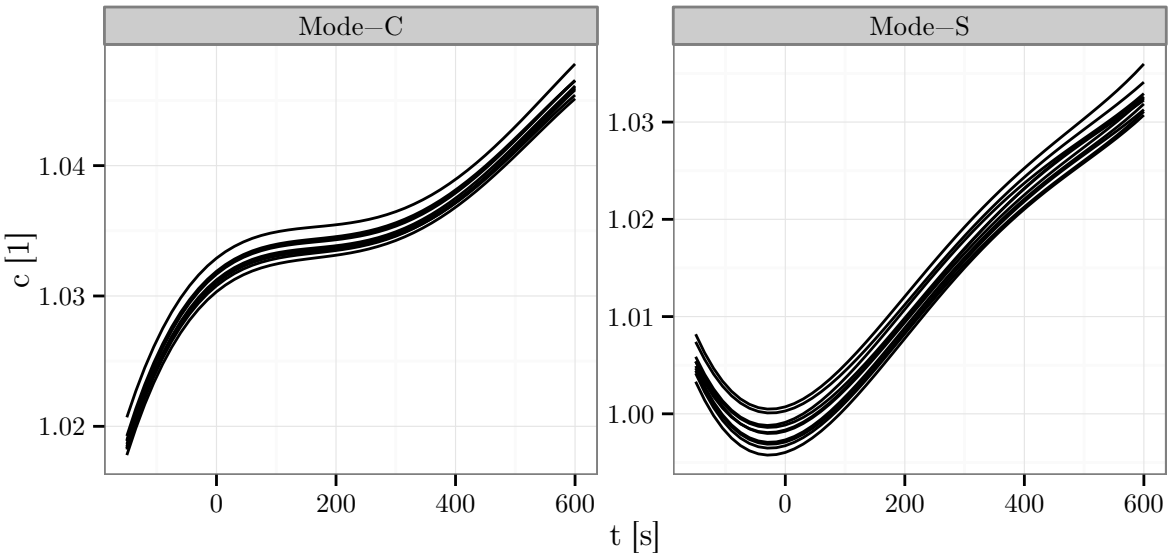


FIGURE 6.7 – Cette figure présente les commandes c_t. 10 commandes de poussée sont tracées correspondant aux 10 plis de la validation croisée.

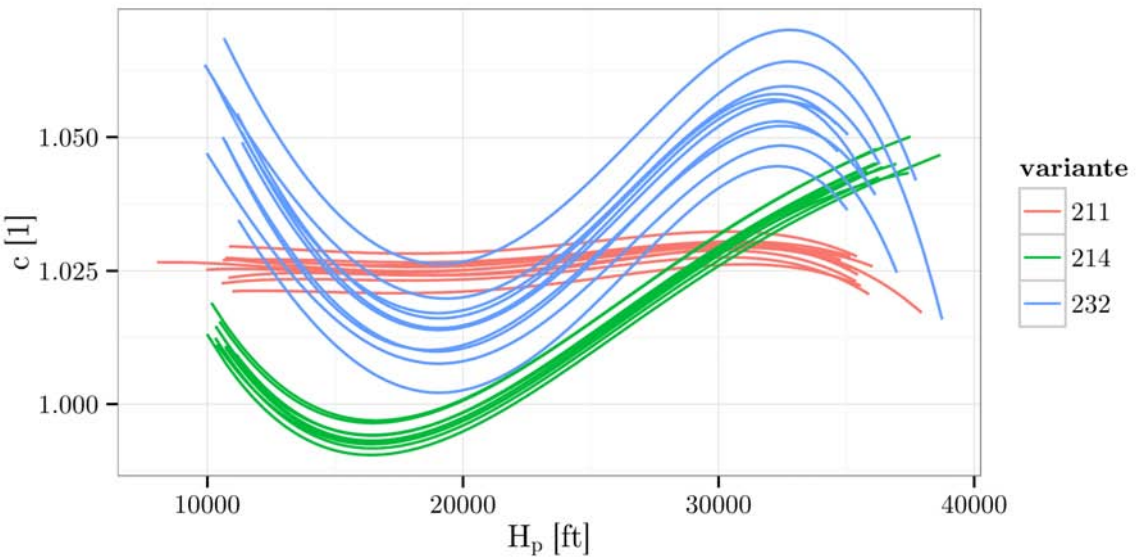


FIGURE 6.8 – Cette figure présente les commandes $c_{variante}[c_{H_p}]$. Les commandes de couleurs différentes sont associées à des variantes d'A320 distinctes. Pour chaque variante, 10 commandes de poussée sont tracées correspondant aux 10 plis de la validation croisée.

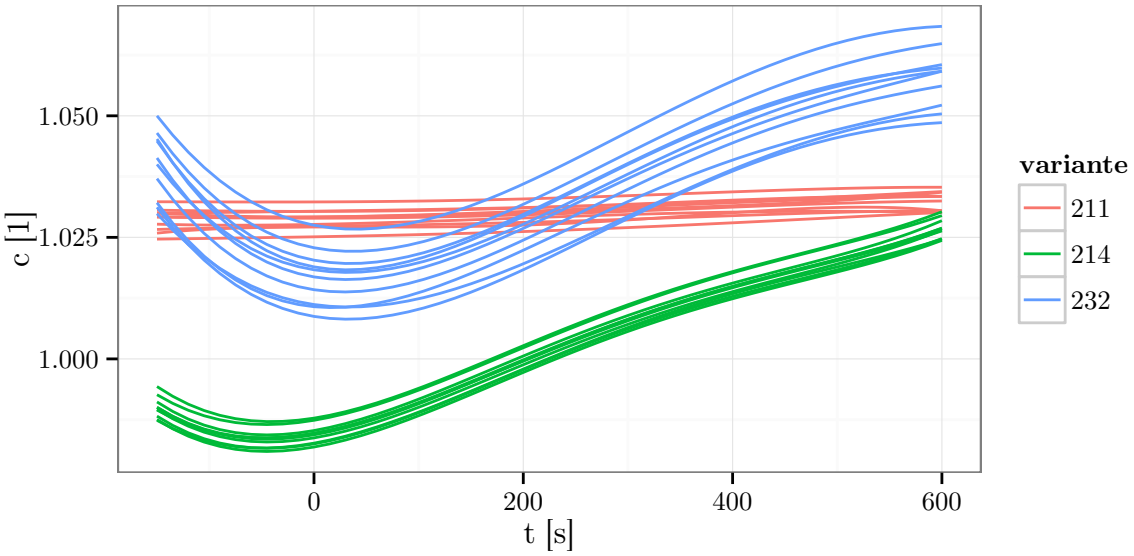


FIGURE 6.9 – Cette figure présente les commandes $c_{variante}[c_t]$. Les commandes de couleurs différentes sont associées à des variantes d'A320 distinctes. Pour chaque variante, 10 commandes de poussée sont tracées correspondant aux 10 plis de la validation croisée.

Seule la variante 232 est modélisée dans la version 3.9 de BADA. Ainsi, la modélisation des forces faite dans BADA ne sera probablement pas adaptée aux autres variantes. Pourtant, c'est ce modèle qui est utilisé indifféremment pour toutes les variantes d'A320. Même si on obtient des commandes de poussée différentes selon les variantes, on ne peut pas conclure que les pilotes appliquent une commande différente suivant la variante de l'appareil. En effet, l'erreur sur l'excès de puissance spécifique modélisé provient de l'erreur sur la modélisation des forces et de l'erreur sur la commande de poussée. Dans notre méthode, on considère que l'erreur sur l'excès de puissance spécifique provient uniquement d'une erreur sur la commande de poussée. Ainsi, la commande apprise va corriger pour partie les erreurs de modélisation des forces.

Évaluation du gain apporté

Avec les trajectoires des *ensembles de validation*, on remplace l'hypothèse *max climb* par la commande de poussée apprise sur l'*ensemble d'apprentissage*. Cette nouvelle hypothèse sur la commande de poussée est utilisée aussi bien pour l'estimation de la masse sur les points passés que pour le calcul de la prédiction de la trajectoire future. Les tableaux 6.8 et 6.9 présentent les résultats obtenus en utilisant le profil de vitesse air V_a observé.

Par rapport à une commande *max climb*, l'utilisation de la commande de poussée apprise permet de réduire la RMSE sur l'altitude, sauf pour la méthode adaptative sur les trajectoires Mode-C. La méthode adaptative sous-estime la masse nécessaire, en particulier

pour les trajectoires Mode-C. La raison de ce biais n'est pas connue. Ceci est d'autant plus étonnant que les méthodes moindres carrés et robuste n'exhibent pas un tel biais.

Avec les méthodes robuste et moindres carrés, l'utilisation des commandes c_t et c_{H_p} permettent de réduire la RMSE de 20 % à 30 % par rapport à une commande *max climb*. L'utilisation des commandes $c_{variante}[c_t]$ et $c_{variante}[c_{H_p}]$ permettent de réduire la RMSE de 35 %.

Sur les données Mode-S, la méthode adaptative a une RMSE similaire à la méthode des moindres carrés. On note toutefois que l'erreur maximale est bien plus faible pour la méthode adaptative. Les erreurs les plus grandes avec les méthodes moindres carrés et robuste sont les cas où la masse $\hat{m}_{11}$ est très grande, supérieure à 90 000 kg. Ainsi, l'erreur maximale peut être réduite en bornant la masse estimée comme c'est le cas pour la méthode adaptative.

TABLE 6.8 – Statistiques, en pieds, sur l'écart entre les trajectoires prédites et les plots observés pour les trajectoires Mode-C. Ces statistiques sont calculées sur l'ensemble des valeurs $\left(H_p^{(pred)}(\hat{m}_{11}) - H_p^{(obs)}\right)(t = 600 \text{ s})$. Les trajectoires sont calculées avec $V_a(t) = V_a{}^{(obs)}(t)$.

méthode	commande	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	*max climb*	-82.4	1479	1168	1481	5493
naïve	*max climb*	-524	651	628	836	6874
adaptative	*max climb*	-303	685	582	749	5535
moindres carrés	*max climb*	-532	653	631	843	6033
robuste	*max climb*	-503	595	603	779	6843
adaptative	c_t	395	702	635	806	6217
moindres carrés	c_t	-3.48	661	460	661	5696
robuste	c_t	23.1	605	439	606	6303
adaptative	c_{H_p}	440	701	653	828	6598
moindres carrés	c_{H_p}	-13.6	671	468	671	5876
robuste	c_{H_p}	13.3	614	445	614	6429

Différences avec l'apprentissage de la masse

Dans la sous-section 6.2, on apprend un modèle reliant les variables disponibles à la masse estimée sur les points futurs. Comme la séparation entre les points futurs et points passés opèrent toujours à 18 000 ft dans nos jeux d'exemples, le modèle est spécialisé pour la prédiction à partir de cette altitude. L'amélioration obtenue provient uniquement d'une masse prédite plus proche de la masse minimisant $e^{[11:51]}$. L'apprentissage de la commande va remettre en cause l'hypothèse *max climb*. Elle va ainsi modifier les fonctions d'erreur $e^{[1:11]}$ et $e^{[11:51]}$. L'amélioration provient d'un minimum plus faible pour cette nouvelle erreur $e^{[11:51]}$ et des masses minimisant $e^{[1:11]}$ et $e^{[11:51]}$ plus proches l'une de l'autre. Les tableaux 6.10 et 6.11 quantifient ces deux aspects. Avec ces deux tableaux, on remarque

TABLE 6.9 – Statistiques, en pieds, sur l'écart entre les trajectoires prédites et les plots observés pour les trajectoires Mode-S. Ces statistiques sont calculées sur l'ensemble des valeurs $\left(H_p^{(pred)}(\hat{m}_{11}) - H_p^{(obs)}\right)(t = 600 \text{ s})$. Les trajectoires sont calculées avec $V_a(t) = V_a{}^{(obs)}(t)$.

méthode	commande	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	*max climb*	-810	1133	1079	1392	5531
naïve	*max climb*	-598	602	673	849	6574
adaptative	*max climb*	-551	554	651	781	4171
moindres carrés	*max climb*	-584	606	663	842	7150
robuste	*max climb*	-571	550	645	792	6371
adaptative	c_t	230	578	459	622	4983
moindres carrés	c_t	19.7	619	396	619	6629
robuste	c_t	33.7	562	378	562	5835
adaptative	c_{H_p}	288	584	480	651	5173
moindres carrés	c_{H_p}	22.4	629	395	629	6897
robuste	c_{H_p}	37.3	569	376	570	6080
adaptative	$c_{variante}[c_t]$	241	513	412	566	4422
moindres carrés	$c_{variante}[c_t]$	21.3	576	335	576	6503
robuste	$c_{variante}[c_t]$	34.4	514	318	515	5701
adaptative	$c_{variante}[c_{H_p}]$	286	511	431	585	4446
moindres carrés	$c_{variante}[c_{H_p}]$	23.6	582	330	583	6816
robuste	$c_{variante}[c_{H_p}]$	40	515	307	517	5988

que pour les trajectoires Mode-C l'erreur $e^{[11:51]}(\hat{m}_{11,futur})$ n'est pas réduite, le gain est principalement dû à l'« alignement » des minima de $e^{[1:11]}$ et $e^{[11:51]}$. Pour les trajectoires Mode-S, on observe une amélioration de l'erreur $e^{[11:51]}(\hat{m}_{11,futur})$ et de l'écart entre les minima de $e^{[1:11]}$ et $e^{[11:51]}$.

TABLE 6.10 – Statistiques, en W/kg, sur l'erreur $e^{[11:51]}$ $(\hat{m}_{11,futur})$. Différentes commandes de poussée sont testées.

trajectoires	commande	moy	écart-type	moy abs	rmse	max abs
Mode-C	*max climb*	12.2	4.02	12.2	12.8	38.5
Mode-C	c_t	12.1	4.04	12.1	12.7	38.5
Mode-C	c_{H_p}	12.1	4.04	12.1	12.7	38.5
Mode-S	*max climb*	7.67	4.49	7.67	8.89	35.6
Mode-S	c_t	7	4.53	7	8.33	34.7
Mode-S	c_{H_p}	6.96	4.53	6.96	8.31	34.7
Mode-S	$c_{variante}[c_t]$	6.86	4.57	6.86	8.24	34.3
Mode-S	$c_{variante}[c_{H_p}]$	6.83	4.56	6.83	8.21	34.2

TABLE 6.11 – Statistiques, en kg, sur l'erreur $\hat{m}_{11} - \hat{m}_{11,futur}$. Différentes commandes de poussée sont testées. La masse $\hat{m}_{11}$ est estimée par la méthode des moindres carrés.

trajectoires	commande	moy	écart-type	moy abs	rmse	max abs
Mode-C	*max climb*	1552	2034	1868	2558	23335
Mode-C	c_t	16.8	2141	1433	2141	22837
Mode-C	c_{H_p}	95.6	2129	1427	2131	23256
Mode-S	*max climb*	1395	1887	1716	2346	24791
Mode-S	c_t	-54.9	1924	1134	1924	23596
Mode-S	c_{H_p}	-39	1877	1095	1877	23492
Mode-S	$c_{variante}[c_t]$	-62.4	1849	1003	1849	22922
Mode-S	$c_{variante}[c_{H_p}]$	-52.5	1788	947	1788	22853

Dans l'apprentissage de la commande, les points passés et points futurs sont traités de manière indifférenciée. Ainsi, la commande obtenue n'est *a priori* pas spécialisée pour la prédiction à partir d'une altitude préalablement fixée. Toutefois, suivant les variables utilisées, l'utilisation de la commande peut être restreinte à un contexte spécifique. Par exemple, la commande c_t utilise la différence entre la date courante et la date à laquelle l'avion est à 18 000 ft. Pour pouvoir utiliser cette commande il faut donc connaître cette date. À contrario, la commande c_{H_p} peut être calculée directement en connaissant l'altitude H_p. Elle peut donc être utilisée par exemple pour prédire la trajectoire à partir de l'altitude $H_p = 15\,000$ ft. Par contre, en prenant toujours 10 points passés, on risque d'avoir des altitudes bien inférieures à celles vues dans les exemples ayant servis à l'apprentissage du profil. Ce problème peut être corrigé en utilisant tous les points de chaque trajectoire en phase de montée.

Conclusion et perspectives

En connaissant le profil de vitesse, apprendre une commande de poussée permet de réduire d'au plus 35 % la RMSE sur l'altitude finale par rapport à l'hypothèse *max climb*. Contrairement à l'apprentissage de la masse future réalisée dans la sous-section 6.2, une même commande de poussée peut être utilisée pour prédire à partir d'altitudes différentes. Les commandes modélisées dans cette section sont relativement simples. La méthode décrite dans cette section peut être employée pour apprendre une commande plus complexe qui serait par exemple modélisée avec un réseau de neurones.

Les masses $\hat{m}_{11,futur}$ sont changées avec ces nouvelles commandes, en apprenant ces nouvelles masses, il est possible que l'on trouve de meilleurs résultats que ceux décrits dans la sous-section 6.2 .

Avec les commandes $c_{variante}$, les trajectoires Mode-S ont été partitionnées selon un critère choisi *a priori*, la variante de l'appareil ici. D'autres facteurs peuvent être pris en compte, comme la compagnie à laquelle l'avion appartient ou l'aéroport de destination par exemple. Il est délicat de décider *a priori* quels facteurs sont pertinents. Pour passer outre ce problème, on pourrait utiliser nos jeux de trajectoires pour construire *a posteriori* une partition des trajectoires, chaque partie ayant en commun une commande de poussée. Ce partitionnement pourrait, par exemple, être construit avec un algorithme des k-moyennes ([MacQueen 67]) dont chaque centroïde i est une commande C_i et chaque point j est une trajectoire j. La « distance » entre le centroïde i et le point j pourrait être choisie comme égale à $\mathcal{E}_j\left(C_i, m_j^*\right)$. La méthode décrite dans cette section servirait à construire le centroïde i à partir des trajectoires de la partie i. Ensuite, pour prédire le futur d'une trajectoire en ayant uniquement ses points passés, il faut pouvoir associer un centroïde i à cette trajectoire. Ceci peut être fait, par exemple, en regardant les « distances » entre un centroïde et la partie passé de la trajectoire.

6.4 Apprendre le profil de vitesse $(cas, Mach)$

Dans la sous-section précédente, on a réduit l'erreur faite sur l'excès de puissance spécifique. L'impact de cette réduction sur l'erreur concernant l'altitude a été évalué en supposant connu le profil de vitesse utilisé par le pilote. Toutefois, dans notre contexte, ce profil n'est pas connu. Il nous faut donc en choisir un à partir des informations disponibles pour pouvoir calculer la trajectoire prédite.

Les profils de vitesse de référence spécifiés par BADA sont paramétrés par un couple de valeurs $(cas, Mach)$. Ces deux valeurs caractérisent un profil de vitesse $V_{acible} = f_{(cas,Mach)}(H_p, T)$ tel que décrit dans la sous-section 2.5.2.

6.4.1 Ajustement du couple $(cas, Mach)$

Les valeurs $(cas, Mach)$ spécifiées dans BADA sont des valeurs par défaut. Pour pouvoir mieux analyser les profils de vitesse réellement suivis par les pilotes, un couple $(cas, Mach)$ a été ajusté sur chaque trajectoire. Ce couple minimise la RMSE ${e^{[11:51]}}_{vitesse}\,(cas, Mach)$

définie par l'équation (6.12). Celle-ci est calculée à partir de la différence entre la vitesse observée V_a et la vitesse modélisée par $f_{(cas,Mach)}$.

$$e^{[i:j]}{}_{vitesse}\,(cas, Mach) = \sqrt{\frac{1}{j-i+1}\sum_{k=i}^{j}\left(f_{(cas,Mach)}(H_{p_k}, T_k) - V_{ak}\right)^2} \tag{6.12}$$

Un profil paramétré par un couple $(cas, Mach)$ possède deux phases. Une phase où la vitesse air est à CAS constante égale à cas. Cette contrainte est maintenue jusqu'à l'altitude de conjonction $H_{p,trans}\,(cas, Mach)$. Au-dessus, l'avion évolue à Mach constant égal à $Mach$. Les 11 points passés de la trajectoire ne sont pas tous à CAS constante ; en effet, la vitesse est limitée à 250 kts en dessous de 10 000 ft. Il y a donc une nécessaire phase d'accélération jusqu'à atteindre le segment CAS constant. En ce qui concerne la phase Mach constant, l'altitude de conjonction est en moyenne à 28 000 ft. Toutefois, l'altitude de conjonction n'est pas atteinte au bout de 10 minutes de montée pour certaines trajectoires. Pour ces trajectoires, une infinité de couple $(cas, Mach)$ minimisent $e^{[i:j]}{}_{vitesse}\,(cas, Mach)$. En effet, si $Mach$ est suffisamment grand, il n'a plus d'influence sur le profil de vitesse dans la plage d'altitude d'intérêt. Ceci nous conduit à remplacer la valeur $Mach$ par NA pour « Not Available ». L'annexe B décrit en détails comment sont obtenus les couples $(cas, Mach)$ minimisant $e^{[i:j]}{}_{vitesse}$.

Analyse des résultats

Les figures 6.10 et 6.11 présentent des exemples de profil de vitesse que l'on peut obtenir avec cet ajustement. Sur ces figures on remarque visuellement que l'erreur $e^{[i:j]}{}_{vitesse}(cas, Mach)$ obtenue est plus petite pour les trajectoires Mode-S que les trajectoires Mode-C. Ceci est attendu car la vitesse V_a pour nos données Mode-S est celle mesurée à bord de de l'avion.

Dans ces deux jeux de trajectoires, toutes les trajectoires ne suivent pas un profil de type $(cas, Mach)$. Les profils « 75% » et « 100% » des figures 6.10 et 6.11 sont constitués de l'enchaînement de deux portions à CAS constante. Dans la modélisation BADA des profils de vitesse, l'avion n'a qu'un seul segment à CAS constante.

Les trajectoires « 100% » sont des trajectoires pour lesquelles le segment Mach constant n'est pas observé. Ceci arrive pour un nombre limité de trajectoires, moins de 8 % pour les trajectoires Mode-C et moins de 5 % pour les trajectoires Mode-S. Si on se restreint aux 50 % des trajectoires les mieux ajustées ces deux chiffres tombent à 2 % et 0,3 % respectivement. De même, on peut douter de certaines valeurs $Mach$ car le nombre de points constituant le segment Mach constant peut être faible. Ce cas est illustré par la trajectoire « 75% » de la figure 6.10. Ceci est probablement limité car la différence entre l'altitude de conjonction et l'altitude du dernier point est supérieur à 2 000 ft dans 75 % des cas. Par rapport à ces deux réserves émises, avoir les points suivants aurait bénéficié à l'estimation du $Mach$. Ceci n'a pas été fait faute de temps.

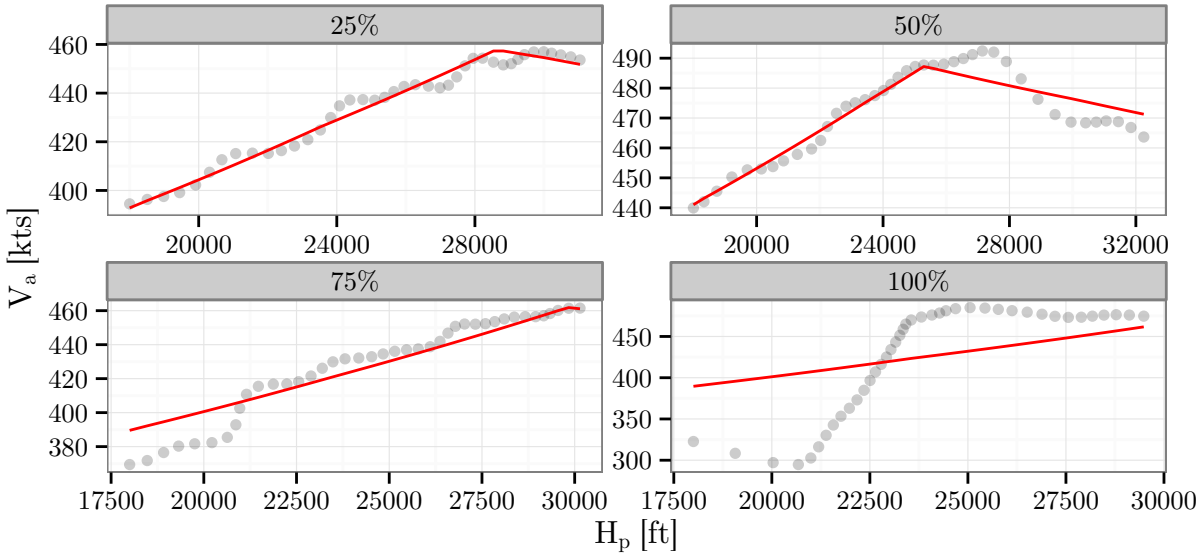


FIGURE 6.10 – Cette figure représente 4 profils de vitesse de trajectoires Mode-C. Les points sont les vitesses V_a et la ligne rouge correspond au profil de vitesse $(cas, Mach)$ obtenu après ajustement du couple $(cas, Mach)$. Le pourcentage indexant chaque trajectoire représente le pourcentage de trajectoires ayant une $e^{[11:51]}{}_{vitesse}\,(cas, Mach)$ inférieure à celle de la trajectoire indexée. Par exemple, 25 % des trajectoires ont une erreur d'ajustement inférieure à la trajectoire « 25% ».

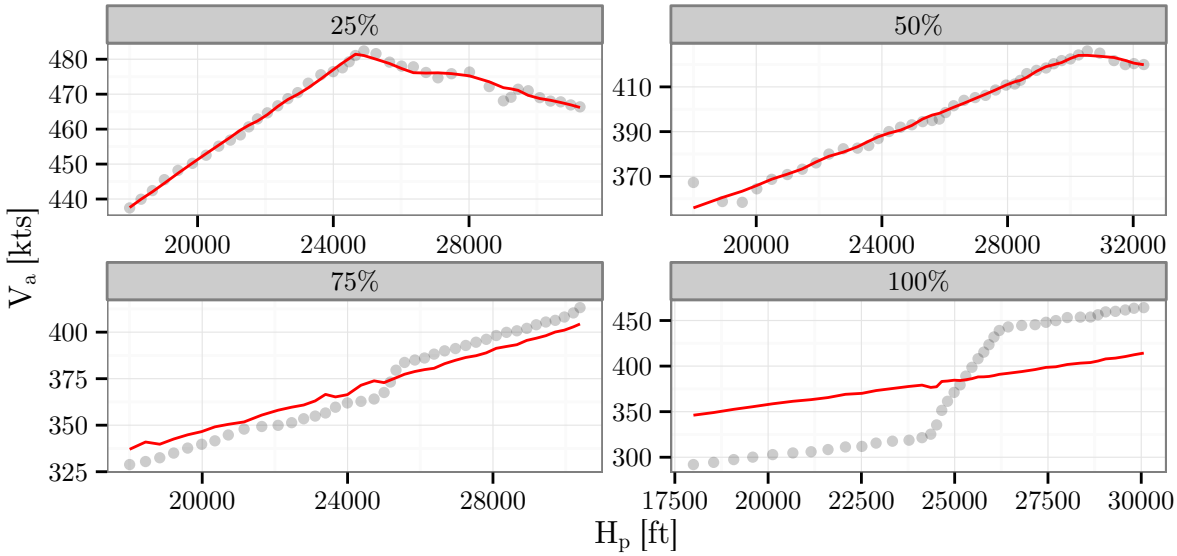


FIGURE 6.11 – Cette figure représente 4 profils de vitesse de trajectoires Mode-S. Les points sont les vitesses V_a et la ligne rouge correspond au profil de vitesse $(cas, Mach)$ obtenu après ajustement du couple $(cas, Mach)$. Le pourcentage indexant chaque trajectoire représente le pourcentage de trajectoires ayant une ${e^{[11:51]}}_{vitesse}\,(cas, Mach)$ inférieure à celle de la trajectoire indexée. Par exemple, 25 % des trajectoires ont une erreur d'ajustement inférieure à la trajectoire « 25% ».

6.4.2 Apprentissage direct des cas et $Mach$ ajustés

Le couple $(cas, Mach)$ ajusté pour chaque trajectoire est inconnu lors du calcul de la prédiction. Ces valeurs ont été ajustées sur les points futurs de la trajectoire. On peut espérer construire un modèle donnant les valeurs de cas et de $Mach$ en fonction d'informations disponibles dans le contexte de la prédiction de trajectoires. Cette idée a déjà été mise en œuvre dans la sous-section 6.2 pour apprendre la masse $\hat{m}_{11,futur}$ qui est aussi estimée sur les points futurs.

Pour ce faire, on considère deux problèmes d'apprentissage complètement séparés, un pour chaque paramètre du couple $(cas, Mach)$. On applique une méthodologie identique à celle décrite dans la sous-section 6.2 pour chacun de ces deux problèmes d'apprentissage.

Analyse des résultats

Les tableaux 6.12 et 6.13 résume les résultats obtenus pour l'apprentissage de la cas avec différents jeux de variables et méthodes. Ces méthodes d'apprentissage sont comparées à des méthodes plus naïves. Parmi ces méthodes, prendre la CAS observée à $t = 0$ réduit la RMSE sur la cas prédite d'au moins 30 % par rapport à l'emploi de la cas moyenne observée. Parmi les méthodes de régression, les modèles linéaires ont des résultats similaires. Les meilleurs résultats sont obtenus avec les méthodes NNet et GBM. Ces méthodes permettent de réduire la RMSE d'au moins 50 % par rapport à l'emploi de la cas moyenne observée.

De même, les tableaux 6.14 et 6.15 présentent les résultats obtenus sur le $Mach$. Dans nos jeux de données aucun avion n'est dans le segment Mach constant à 18 000 ft. Ainsi, utiliser le $Mach$ à $t = 0$ n'est pas une bonne idée. Le $Mach_{BADA}$ est similaire au $Mach$ moyen pour les données Mode-C. Parmi les méthodes de régression, les modèles linéaires ont des résultats similaires. Les meilleurs résultats sont obtenus avec les méthodes NNet et GBM. Ces méthodes permettent de réduire la RMSE d'au plus 15 % par rapport à l'emploi du $Mach$ moyen observé. Le gain obtenu est ainsi très limité. Ceci n'est pas surprenant car le $Mach$ ne joue *a priori* aucun rôle dans la trajectoire passée contrairement à la masse et la cas. En effet, la masse intervient dans le calcul de l'excès de puissance spécifique qui est une grandeur observable sur les points passés au travers de la variation d'énergie spécifique. De même, la trajectoire passée peut nous renseigner sur la cas choisie car certaines trajectoires sont déjà sur le segment CAS constante à 18 000 ft.

Les tableaux 6.16 et 6.17 présentent les erreurs obtenues sur la vitesse V_a en utilisant les valeurs de cas et $Mach$ prédites. Ces erreurs sont calculées sur les points futurs de la trajectoire à altitude pression H_p correspondante. Par rapport à l'emploi du cas et $Mach$ moyen, l'utilisation de la cas à $t = 0$ permet de réduire la RMSE sur la vitesse V_a de 20 %. L'utilisation des modèles appris permet de réduire cette RMSE jusqu'à 35 %. Les statistiques présentées dans ces tableaux ne rendent compte que de l'erreur entre V_a et V_{acible}. Il faut toutefois avoir conscience qu'avec la modélisation du profil de vitesse, l'erreur ne peut descendre en dessous d'une certaine valeur qui correspond à la ligne $f_{(cas,Mach)}$ des tableaux. La figure 6.12 permet de mieux se rendre compte de la décomposition de l'erreur entre l'erreur de prédiction du couple $(cas, Mach)$ et l'erreur due à la modélisation du

TABLE 6.12 – Statistiques en nœuds pour les trajectoires Mode-C sur la différence entre la *cas* ajustée sur les points futurs et la *cas* choisie à partir des informations disponibles au moment de la prédiction.

variables	méthode	moy	écart-type	moy abs	rmse	max abs
-	$\overline{cas}$	6.46e-15	19	15.1	19	64
-	cas_{BADA}	1.23	19	15.4	19.1	65.2
-	$cas\,(t=0)$	-3.24	12	7.31	12.4	85.1
r	$cas_{\text{MLR-FS}}$	0.0382	9.2	6.01	9.2	50.9
r	cas_{Ridge}	0.0125	9.22	6.09	9.22	49.9
r	cas_{PCR}	0.0167	9.53	6.53	9.53	61.9
r	cas_{NNet}	0.0121	9.15	5.93	9.14	47
r	cas_{GBM}	0.0161	9.12	5.82	9.12	53.9
$\hat{m}r$	$cas_{\text{MLR-FS}}$	0.05	9.17	5.98	9.17	46.7
$\hat{m}r$	cas_{Ridge}	-0.0075	9.32	6.17	9.32	59.6
$\hat{m}r$	cas_{PCR}	-0.0127	9.36	6.29	9.36	52
$\hat{m}r$	cas_{NNet}	-0.0158	9.2	5.98	9.2	46.4
$\hat{m}r$	cas_{GBM}	0.0211	9.13	5.82	9.13	54.9
$\hat{m}rw$	$cas_{\text{MLR-FS}}$	-0.0447	9.14	5.98	9.14	49.8
$\hat{m}rw$	cas_{Ridge}	-0.0105	9.3	6.16	9.3	59.3
$\hat{m}rw$	cas_{PCR}	-0.0147	9.5	6.47	9.49	51.8
$\hat{m}rw$	cas_{NNet}	0.0341	9.18	6	9.18	46.1
$\hat{m}rw$	cas_{GBM}	0.00439	9.07	5.79	9.07	51.5
$\hat{m}prw$	$cas_{\text{MLR-FS}}$	-0.00277	9.18	6.03	9.18	49.6
$\hat{m}prw$	cas_{Ridge}	0.0211	9.18	6.13	9.18	47
$\hat{m}prw$	cas_{PCR}	-0.0142	9.43	6.45	9.43	49.4
$\hat{m}prw$	cas_{NNet}	0.0354	9.09	5.92	9.09	46.1
$\hat{m}prw$	cas_{GBM}	0.0082	8.83	5.63	8.83	51.8
$c\hat{m}prw$	cas_{GBM}	0.00986	8.64	5.47	8.64	50.3

TABLE 6.13 – Statistiques en nœuds pour les trajectoires Mode-S sur la différence entre la *cas* ajustée sur les points futurs et la *cas* choisie à partir des informations disponibles au moment de la prédiction.

variables	méthode	moy	écart-type	moy abs	rmse	max abs
-	$\overline{cas}$	-2.15e-14	17.2	12.8	17.2	71
-	cas_{BADA}	14.4	17.2	19.9	22.4	85.4
-	$cas\,(t=0)$	-2.15	11.8	5.89	12	90
r	$cas_{\text{MLR-FS}}$	-0.00918	9.32	6.28	9.31	70.9
r	cas_{Ridge}	0.0159	9.66	6.46	9.65	73.1
r	cas_{PCR}	0.0167	9.36	6.38	9.35	68.1
r	cas_{NNet}	0.0225	8.6	5.19	8.59	79.6
r	cas_{GBM}	0.0212	8.64	4.99	8.64	86.8
$\hat{m}r$	$cas_{\text{MLR-FS}}$	-0.0285	9.02	5.95	9.01	75.8
$\hat{m}r$	cas_{Ridge}	0.048	9.07	5.9	9.06	74.4
$\hat{m}r$	cas_{PCR}	0.0516	8.94	5.94	8.93	78
$\hat{m}r$	cas_{NNet}	-0.00143	8.46	5.14	8.46	80.9
$\hat{m}r$	cas_{GBM}	0.00874	8.6	4.97	8.6	86.5
$\hat{m}pr$	$cas_{\text{MLR-FS}}$	0.0216	9.16	6.08	9.15	76.7
$\hat{m}pr$	cas_{Ridge}	0.0569	9.05	5.91	9.04	74.4
$\hat{m}pr$	cas_{PCR}	0.0563	8.91	5.94	8.9	78.2
$\hat{m}pr$	cas_{NNet}	0.0295	8.4	5.11	8.4	80.9
$\hat{m}pr$	cas_{GBM}	-0.0205	8.48	4.82	8.48	83.8
$c\hat{m}pr$	cas_{GBM}	-0.013	8.55	4.93	8.54	84
$c\hat{m}prt$	cas_{GBM}	-0.045	8.66	5.03	8.65	83.8

TABLE 6.14 – Statistiques pour les trajectoires Mode-C, sans dimensions, sur la différence entre le $Mach$ ajusté sur les points futurs et le $Mach$ choisi à partir des informations disponibles au moment de la prédiction.

variables	méthode	moy	écart-type	moy abs	rmse	max abs
-	$\overline{Mach}$	3.7e-17	0.0216	0.0153	0.0216	0.126
-	$Mach_{BADA}$	-0.00286	0.0216	0.016	0.0218	0.123
r	$Mach_{\text{MLR-FS}}$	3.02e-05	0.0198	0.0138	0.0198	0.123
r	$Mach_{\text{Ridge}}$	-2.32e-06	0.0198	0.0138	0.0198	0.124
r	$Mach_{\text{PCR}}$	4.45e-05	0.0199	0.0138	0.0199	0.123
r	$Mach_{\text{NNet}}$	-4.25e-05	0.0196	0.0136	0.0196	0.126
r	$Mach_{\text{GBM}}$	-6.88e-05	0.0193	0.0133	0.0193	0.125
$\hat{m}r$	$Mach_{\text{MLR-FS}}$	3.18e-05	0.0198	0.0138	0.0198	0.123
$\hat{m}r$	$Mach_{\text{Ridge}}$	-9.74e-07	0.0198	0.0138	0.0198	0.123
$\hat{m}r$	$Mach_{\text{PCR}}$	1.02e-05	0.0199	0.0138	0.0199	0.123
$\hat{m}r$	$Mach_{\text{NNet}}$	2.63e-05	0.0196	0.0136	0.0196	0.126
$\hat{m}r$	$Mach_{\text{GBM}}$	-4.74e-05	0.0192	0.0132	0.0192	0.125
$\hat{m}rw$	$Mach_{\text{MLR-FS}}$	-3.8e-05	0.0199	0.0138	0.0199	0.124
$\hat{m}rw$	$Mach_{\text{Ridge}}$	-3.48e-06	0.0198	0.0137	0.0197	0.122
$\hat{m}rw$	$Mach_{\text{PCR}}$	-5.02e-05	0.0199	0.0138	0.0199	0.126
$\hat{m}rw$	$Mach_{\text{NNet}}$	-5.28e-06	0.0195	0.0135	0.0195	0.124
$\hat{m}rw$	$Mach_{\text{GBM}}$	-0.000142	0.0191	0.0132	0.0191	0.127
$\hat{m}prw$	$Mach_{\text{MLR-FS}}$	-6.9e-05	0.0198	0.0137	0.0198	0.122
$\hat{m}prw$	$Mach_{\text{Ridge}}$	6.09e-06	0.0196	0.0137	0.0196	0.124
$\hat{m}prw$	$Mach_{\text{PCR}}$	-2.97e-05	0.0197	0.0137	0.0197	0.128
$\hat{m}prw$	$Mach_{\text{NNet}}$	2.87e-05	0.0194	0.0134	0.0194	0.127
$\hat{m}prw$	$Mach_{\text{GBM}}$	-8.82e-05	0.019	0.0131	0.019	0.126
$c\hat{m}prw$	$Mach_{\text{GBM}}$	-0.00011	0.0189	0.013	0.0189	0.127

TABLE 6.15 – Statistiques pour les trajectoires Mode-S, sans dimensions, sur la différence entre le $Mach$ ajusté sur les points futurs et le $Mach$ choisi à partir des informations disponibles au moment de la prédiction.

variables	méthode	moy	écart-type	moy abs	rmse	max abs
-	$\overline{Mach}$	-4.5e-17	0.0225	0.0169	0.0225	0.135
-	$Mach_{BADA}$	0.00944	0.0225	0.0161	0.0244	0.145
r	$Mach_{\text{MLR-FS}}$	-1.87e-05	0.02	0.0144	0.02	0.123
r	$Mach_{\text{Ridge}}$	0.0001	0.0204	0.0148	0.0204	0.115
r	$Mach_{\text{PCR}}$	1.23e-05	0.0199	0.0144	0.0199	0.117
r	$Mach_{\text{NNet}}$	-1.32e-05	0.0197	0.0142	0.0197	0.12
r	$Mach_{\text{GBM}}$	-0.000126	0.0197	0.0139	0.0197	0.131
$\hat{m}r$	$Mach_{\text{MLR-FS}}$	-5.01e-05	0.0201	0.0145	0.0201	0.123
$\hat{m}r$	$Mach_{\text{Ridge}}$	0.00014	0.0205	0.0149	0.0205	0.122
$\hat{m}r$	$Mach_{\text{PCR}}$	0.000141	0.0199	0.0144	0.0199	0.123
$\hat{m}r$	$Mach_{\text{NNet}}$	0.000194	0.0199	0.0143	0.0199	0.122
$\hat{m}r$	$Mach_{\text{GBM}}$	-0.000113	0.0196	0.0138	0.0196	0.128
$\hat{m}pr$	$Mach_{\text{MLR-FS}}$	4.06e-05	0.0196	0.014	0.0196	0.124
$\hat{m}pr$	$Mach_{\text{Ridge}}$	0.000123	0.0197	0.0142	0.0197	0.121
$\hat{m}pr$	$Mach_{\text{PCR}}$	0.00015	0.0194	0.0139	0.0194	0.119
$\hat{m}pr$	$Mach_{\text{NNet}}$	-0.000126	0.0194	0.0138	0.0194	0.126
$\hat{m}pr$	$Mach_{\text{GBM}}$	-0.000165	0.0189	0.013	0.0189	0.134
$c\hat{m}pr$	$Mach_{\text{GBM}}$	-0.000235	0.0189	0.0129	0.0189	0.133
$c\hat{m}prt$	$Mach_{\text{GBM}}$	-9.9e-05	0.0191	0.013	0.0191	0.132

profil de vitesse en profil de type $(cas, Mach)$.

TABLE 6.16 – Statistiques en nœuds sur $(V_{acible} - V_a)$ $(t \geqslant 0)$ pour les trajectoires Mode-C.

variables	méthode	moy	écart-type	moy abs	rmse	max abs
-	$f_{(cas,Mach)}$	0.049	9.29	5.6	9.29	122
-	$f_{\left(\overline{cas},\overline{Mach}\right)}$	1.55	23.7	17.7	23.8	131
-	$f_{\left(cas(t=0),\overline{Mach}\right)}$	-4.29	18.5	11.5	19	157
-	$f_{(cas_{BADA},Mach_{BADA})}$	2.13	23.6	17.9	23.7	133
r	$f_{(cas_{\text{MLR-FS}},Mach_{\text{MLR-FS}})}$	0.676	15.7	10.2	15.7	115
r	$f_{\left(cas_{\text{Ridge}},Mach_{\text{Ridge}}\right)}$	0.66	15.7	10.2	15.7	115
r	$f_{(cas_{\text{PCR}},Mach_{\text{PCR}})}$	0.679	15.9	10.5	15.9	115
r	$f_{(cas_{\text{NNet}},Mach_{\text{NNet}})}$	0.642	15.6	10.1	15.6	111
r	$f_{(cas_{\text{GBM}},Mach_{\text{GBM}})}$	0.643	15.5	9.97	15.5	112
$\hat{m}r$	$f_{(cas_{\text{MLR-FS}},Mach_{\text{MLR-FS}})}$	0.682	15.6	10.1	15.6	116
$\hat{m}r$	$f_{\left(cas_{\text{Ridge}},Mach_{\text{Ridge}}\right)}$	0.629	15.7	10.2	15.8	121
$\hat{m}r$	$f_{(cas_{\text{PCR}},Mach_{\text{PCR}})}$	0.63	15.8	10.3	15.8	116
$\hat{m}r$	$f_{(cas_{\text{NNet}},Mach_{\text{NNet}})}$	0.617	15.6	10.1	15.6	111
$\hat{m}r$	$f_{(cas_{\text{GBM}},Mach_{\text{GBM}})}$	0.653	15.5	9.97	15.5	112
$\hat{m}rw$	$f_{(cas_{\text{MLR-FS}},Mach_{\text{MLR-FS}})}$	0.58	15.6	10.1	15.6	110
$\hat{m}rw$	$f_{\left(cas_{\text{Ridge}},Mach_{\text{Ridge}}\right)}$	0.619	15.7	10.2	15.7	121
$\hat{m}rw$	$f_{(cas_{\text{PCR}},Mach_{\text{PCR}})}$	0.646	15.9	10.5	15.9	115
$\hat{m}rw$	$f_{(cas_{\text{NNet}},Mach_{\text{NNet}})}$	0.661	15.6	10.1	15.6	111
$\hat{m}rw$	$f_{(cas_{\text{GBM}},Mach_{\text{GBM}})}$	0.619	15.5	9.93	15.5	112
$\hat{m}prw$	$f_{(cas_{\text{MLR-FS}},Mach_{\text{MLR-FS}})}$	0.616	15.6	10.1	15.6	115
$\hat{m}prw$	$f_{\left(cas_{\text{Ridge}},Mach_{\text{Ridge}}\right)}$	0.662	15.6	10.2	15.6	115
$\hat{m}prw$	$f_{(cas_{\text{PCR}},Mach_{\text{PCR}})}$	0.646	15.8	10.5	15.8	114
$\hat{m}prw$	$f_{(cas_{\text{NNet}},Mach_{\text{NNet}})}$	0.653	15.5	10	15.5	115
$\hat{m}prw$	$f_{(cas_{\text{GBM}},Mach_{\text{GBM}})}$	0.629	15.2	9.76	15.2	112
$c\hat{m}prw$	$f_{(cas_{\text{GBM}},Mach_{\text{GBM}})}$	0.623	15.1	9.64	15.1	113

Conclusion et limites de cette étude

Le profil de vitesse cible V_{acible}, tel que modélisé dans BADA, est paramétré par le couple $(cas, Mach)$: on a $V_{acible} = f_{(cas,Mach)}$ avec f qui provient d'équations physiques modélisant l'atmosphère. Ainsi, en prédisant $(cas, Mach)$, on cherche à réduire l'erreur $V_a - f_{(cas,Mach)}$. Ce choix impose l'emploi d'une fonction de risque spécifique à notre problème. Dans cette sous-section, le choix a été fait de garder un risque quadratique. En effet, la fonction de risque quadratique est gérée de manière efficace par toutes les méthodes d'apprentissage que l'on a considérées. Pour ce faire, on a choisi d'apprendre séparément la cas et le $Mach$ en se servant de profils de vitesse préalablement ajustés.

TABLE 6.17 – Statistiques en nœuds sur $(V_{acible} - V_a)$ $(t \geqslant 0)$ pour les trajectoires Mode-S.

variables	méthode	moy	écart-type	moy abs	rmse	max abs
-	$f_{(cas,Mach)}$	0.0289	8.32	4.01	8.32	114
-	$f_{\left(\overline{cas},\overline{Mach}\right)}$	1.15	21.8	15.9	21.9	120
-	$f_{\left(cas(t=0),\overline{Mach}\right)}$	-3.43	17.8	9.81	18.1	134
-	$f_{(cas_{BADA},Mach_{BADA})}$	15.6	22.2	21.6	27.1	139
r	$f_{(cas_{\text{MLR-FS}},Mach_{\text{MLR-FS}})}$	0.539	14.9	9.62	15	131
r	$f_{\left(cas_{\text{Ridge}},Mach_{\text{Ridge}}\right)}$	0.511	15.3	9.84	15.3	134
r	$f_{(cas_{\text{PCR}},Mach_{\text{PCR}})}$	0.6	14.9	9.73	15	127
r	$f_{(cas_{\text{NNet}},Mach_{\text{NNet}})}$	0.544	14.3	8.69	14.3	143
r	$f_{(cas_{\text{GBM}},Mach_{\text{GBM}})}$	0.492	14.2	8.46	14.2	143
$\hat{m}r$	$f_{(cas_{\text{MLR-FS}},Mach_{\text{MLR-FS}})}$	0.474	14.6	9.42	14.6	138
$\hat{m}r$	$f_{\left(cas_{\text{Ridge}},Mach_{\text{Ridge}}\right)}$	0.552	14.7	9.41	14.8	136
$\hat{m}r$	$f_{(cas_{\text{PCR}},Mach_{\text{PCR}})}$	0.61	14.6	9.41	14.6	141
$\hat{m}r$	$f_{(cas_{\text{NNet}},Mach_{\text{NNet}})}$	0.61	14.2	8.65	14.2	144
$\hat{m}r$	$f_{(cas_{\text{GBM}},Mach_{\text{GBM}})}$	0.504	14.1	8.41	14.2	143
$\hat{m}pr$	$f_{(cas_{\text{MLR-FS}},Mach_{\text{MLR-FS}})}$	0.45	14.6	9.43	14.6	139
$\hat{m}pr$	$f_{\left(cas_{\text{Ridge}},Mach_{\text{Ridge}}\right)}$	0.508	14.6	9.32	14.6	136
$\hat{m}pr$	$f_{(cas_{\text{PCR}},Mach_{\text{PCR}})}$	0.553	14.4	9.35	14.4	141
$\hat{m}pr$	$f_{(cas_{\text{NNet}},Mach_{\text{NNet}})}$	0.538	14	8.58	14	144
$\hat{m}pr$	$f_{(cas_{\text{GBM}},Mach_{\text{GBM}})}$	0.413	14	8.18	14	144
$c\hat{m}pr$	$f_{(cas_{\text{GBM}},Mach_{\text{GBM}})}$	0.429	14	8.26	14	144
$c\hat{m}prt$	$f_{(cas_{\text{GBM}},Mach_{\text{GBM}})}$	0.43	14.1	8.34	14.1	144

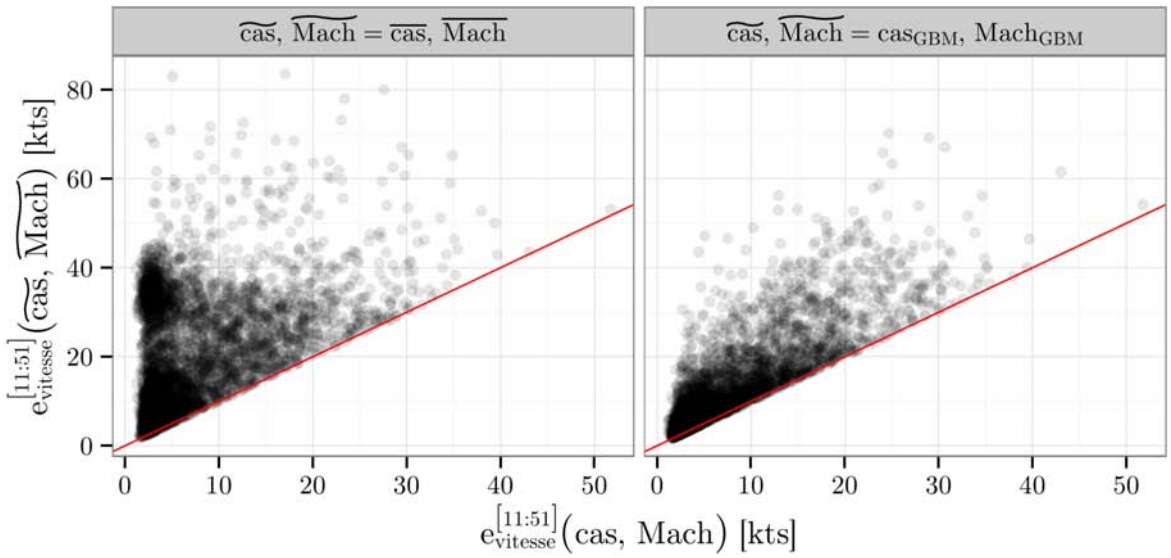


FIGURE 6.12 – Cette figure présente l'erreur $e^{[11:51]}{}_{vitesse}$ de chaque trajectoire en fonction du couple $(\widetilde{cas}, \widetilde{Mach})$ utilisé. Cette erreur est comparée à l'utilisation du couple $(cas, Mach)$ ajusté. Les trajectoires considérées proviennent des données Mode-C. Les données Mode-S donnent une figure similaire. La droite rouge est d'équation $y = x$.

L'application de méthodes d'apprentissage sur ces deux sous problèmes permet de réduire l'écart entre le profil de vitesse prévu et celui effectivement suivi. Ainsi la RMSE entre la vitesse observée et celle prédite est réduite de 35 % comparé à des choix moyens [2].

Plusieurs points peuvent limiter l'amélioration apportée par l'utilisation de méthodes d'apprentissage. En effet, on ne tient pas compte des interactions entre la *cas* prédite et le $Mach$ prédit dans le calcul de V_{acible}. On a traité séparément la *cas* et le $Mach$ sans que f n'intervienne. Outre ce problème d'interaction, avoir une erreur nulle sur le couple $(cas, Mach)$ n'engendre pas la même erreur sur V_a suivant la trajectoire considérée. En effet, la majorité des trajectoires se conforme à un profil de type $(cas, Mach)$ tandis que certaines en sont assez éloignées. Ces deux problèmes sont résolus en considérant une fonction de risque spécifique prenant directement en compte l'erreur $V_a - f_{(cas,Mach)}$.

6.5 Comparaison des différentes méthodes

Dans cette section, des méthodes d'apprentissage sont directement appliquées à l'altitude $H_p\,(t = 600 \text{ s})$. Ces résultats sont ensuite comparés à ceux du modèle physique utilisant les méthodes présentées dans les sections 5, 6.2, 6.3 et 6.4.

6.5.1 Apprendre directement l'altitude

Apprendre directement l'altitude $H_p\,(t = 600 \text{ s})$ permet de se passer complètement d'un modèle physique. Cette approche est illustrée par la figure 6.13. La qualité du modèle appris sera uniquement évaluée en comparant l'altitude prédite à l'altitude observée à $t = 600$ s. Le tableau 6.1 présente tout les algorithmes d'apprentissage testés et la grille d'hyper-paramètres associé suivant l'algorithme. Les algorithmes $TuneGrid$ ou $TuneGridCV$ décrits dans 6.1 permettent de choisir les hyper-paramètres des algorithmes d'apprentissage. Les statistiques présentées sont calculées en se servant des prédictions obtenues par $CV_{10}(TuneGrid\,(\mathcal{A}_\lambda, grid$ ou $CV_{10}(TuneGridCV\,(\mathcal{A}_\lambda, grid)\,, S)$.

Les tableaux 6.18 et 6.19 présentent les résultats des différentes méthodes. La RMSE sur l'altitude prédite diminue avec l'ajout de variables explicatives. La plus grande diminution de RMSE pour les trajectoires Mode-C se fait sur l'ajout des données météorologiques w. Les modèles linéaires appris ont des RMSE comparables. Les modèles GBM et NNet ont des RMSE plus faibles que les modèles linéaires avec un avantage pour le modèle appris via la méthode GBM. À jeu de variables égal, l'écart de RMSE entre GBM et les modèles linéaires n'est toutefois que de 40 ft.

GBM utilisé avec toutes les variables disponibles permet de réduire de moitié la RMSE par rapport à celle obtenue en utilisant le modèle BADA avec les paramètres de références.

2. Dans notre cas, « moyen » signifie que l'on a pris la *cas* moyenne et le $Mach$ moyen. Le profil « moyen » obtenu n'est donc pas la moyenne des profils de vitesse car $E\,[f\,(X)] \neq f\,(E\,[X])$ en général.

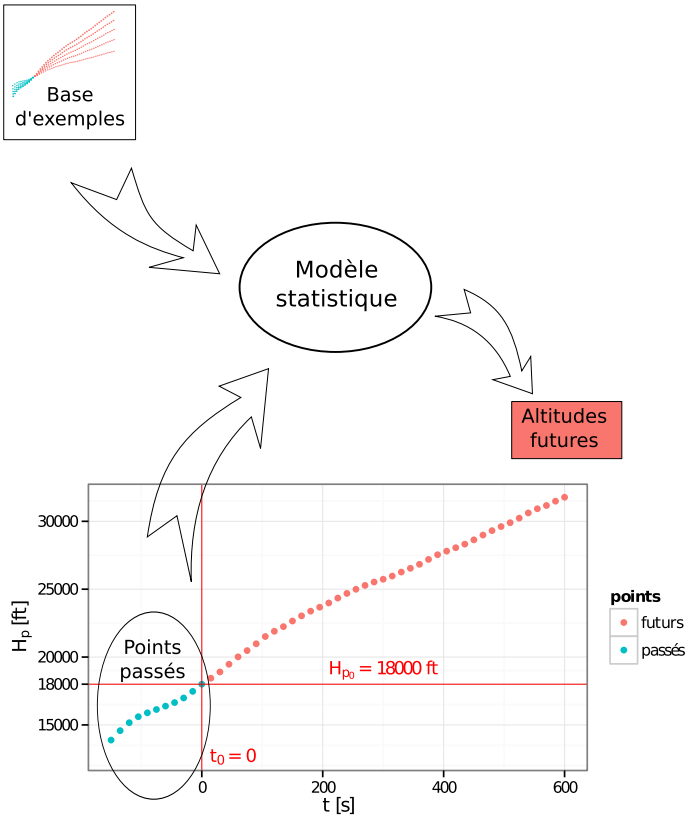


FIGURE 6.13 – Cette figure illustre l'approche consistant à apprendre directement l'altitude finale.

TABLE 6.18 – Statistiques, en pieds, sur l'écart entre l'altitude prédite à 10 minutes et celle observée pour les trajectoires Mode-C. Ces statistiques sont calculées sur l'ensemble des valeurs $H_p^{(pred)} - H_p^{(obs)}$ ($t = 600$ s). Chacune de ces valeurs est associée à une trajectoire.

variables	méthode	moy	écart-type	moy abs	rmse	max abs
$BADA_{ref}$		113	1591	1274	1595	6415
r	MLR-FS	-2.82	995	742	995	7536
r	Ridge	-0.801	991	738	991	7526
r	PCR	-0.853	990	740	990	7516
r	NNet	-9.87	965	722	965	6650
r	GBM	-2.21	948	708	948	7316
$\hat{m}r$	MLR-FS	-5.25	980	734	980	7116
$\hat{m}r$	Ridge	-1.3	983	736	983	7714
$\hat{m}r$	PCR	-1.26	980	736	980	7694
$\hat{m}r$	NNet	-3.36	961	722	961	6843
$\hat{m}r$	GBM	-2.88	940	701	939	7416
$\hat{m}rw$	MLR-FS	-4.09	914	668	914	7963
$\hat{m}rw$	Ridge	-2.03	918	674	918	8068
$\hat{m}rw$	PCR	-2.2	916	674	916	7896
$\hat{m}rw$	NNet	-3.6	903	661	903	7205
$\hat{m}rw$	GBM	-2.43	904	664	904	7562
$\hat{m}prw$	MLR-FS	-2.75	905	665	905	8111
$\hat{m}prw$	Ridge	-1.38	910	671	910	8248
$\hat{m}prw$	PCR	-2.59	909	671	909	8055
$\hat{m}prw$	NNet	-0.848	892	656	892	7207
$\hat{m}prw$	GBM	-1.7	885	653	885	7784
$c\hat{m}prw$	GBM	-2.95	859	633	859	7852

TABLE 6.19 – Statistiques, en pieds, sur l'écart entre l'altitude prédite à 10 minutes et celle observée pour les trajectoires Mode-S. Ces statistiques sont calculées sur l'ensemble des valeurs $H_p^{(pred)} - H_p^{(obs)}$ ($t = 600$ s). Chacune de ces valeurs est associée à une trajectoire.

variables	méthode	moy	écart-type	moy abs	rmse	max abs
$BADA_{ref}$		-1032	1343	1348	1693	6065
r	MLR-FS	-3.25	883	659	883	5309
r	Ridge	-5.16	895	668	895	5954
r	PCR	-4.28	884	660	884	6293
r	NNet	-1.62	881	664	880	5260
r	GBM	1.1	848	639	848	3755
$\hat{m}r$	MLR-FS	-5.31	877	659	877	4869
$\hat{m}r$	Ridge	-0.6	880	664	880	4390
$\hat{m}r$	PCR	-10	885	666	885	5340
$\hat{m}r$	NNet	-0.637	867	660	867	4054
$\hat{m}r$	GBM	-1.28	856	646	856	4438
$\hat{m}pr$	MLR-FS	-5.31	877	659	877	4869
$\hat{m}pr$	Ridge	-0.0828	882	665	881	4437
$\hat{m}pr$	PCR	-9.61	888	669	888	5373
$\hat{m}pr$	NNet	1.37	878	662	878	4271
$\hat{m}pr$	GBM	-1.05	849	638	849	4124
$c\hat{m}pr$	GBM	-2.32	841	632	841	4135
$c\hat{m}prt$	GBM	1.01	830	621	830	4021

6.5.2 Prédiction avec le modèle physique BADA

Pour le modèle physique, des paramètres importants comme la masse, la commande de poussée et le profil de vitesse choisi sont inconnus. Dans tout le chapitre 5, on a construit des modèles permettant d'inférer ces paramètres à partir des informations disponibles lors du calcul de la trajectoire prédite. Ces améliorations ont été évaluées en supposant connus le reste des paramètres. Dans cette section, on se place dans un cadre plus réaliste. Les paramètres utilisés sont uniquement ceux inférés avec les méthodes présentées précédemment.

TABLE 6.20 – Statistiques, en pieds, sur l'écart entre les trajectoires prédites et les plots observés pour les trajectoires Mode-C. Ces statistiques sont calculées sur l'ensemble des valeurs $\left(H_p^{(pred)}(\hat{m}_{11}) - H_p^{(obs)}\right)(t = 600 \text{ s})$. Chacune de ces valeurs est associée à une trajectoire. Différentes commandes de poussée sont testées.

masse	commande	V_a	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	*max climb*	ref	113	1591	1274	1595	6415
adaptative	*max climb*	ref	-76.4	1035	807	1038	7163
moindres carrés	*max climb*	ref	-306	1028	825	1072	6682
robuste	*max climb*	ref	-276	991	803	1029	7566
adaptative	c_t	ref	629	989	919	1172	7667
moindres carrés	c_t	ref	232	985	756	1012	7196
robuste	c_t	ref	259	949	740	984	7369
adaptative	c_{H_p}	ref	684	993	949	1205	8073
moindres carrés	c_{H_p}	ref	229	995	764	1021	7401
robuste	c_{H_p}	ref	256	958	748	992	7561
GBM	*max climb*	ref	166	876	669	891	7020
$mass_{ref}$	*max climb*	GBM	15.7	1660	1318	1660	6402
adaptative	*max climb*	GBM	-174	971	749	986	7376
moindres carrés	*max climb*	GBM	-403	951	774	1033	6851
robuste	*max climb*	GBM	-374	916	749	989	7026
adaptative	c_t	GBM	535	958	843	1097	7761
moindres carrés	c_t	GBM	136	943	690	952	7284
robuste	c_t	GBM	164	907	673	922	7458
adaptative	c_{H_p}	GBM	587	966	872	1131	8166
moindres carrés	c_{H_p}	GBM	131	958	698	967	7487
robuste	c_{H_p}	GBM	159	922	680	935	7652
GBM	*max climb*	GBM	68	824	590	826	7236

Le profil de vitesse utilisé est un profil de type $(cas, Mach)$. On peut choisir le couple de valeurs en utilisant le $(cas, Mach)$ de référence contenu dans les fichiers BADA. On a également testé le modèle GBM appris dans la section 6.4 avec les variables $\hat{cmprt}$ et $\hat{cmprw}$ pour les trajectoires Mode-C et Mode-S respectivement. De même, concernant la

TABLE 6.21 – Statistiques, en pieds, sur l'écart entre les trajectoires prédites et les plots observés pour les trajectoires Mode-S. Ces statistiques sont calculées sur l'ensemble des valeurs $\left(H_p^{(pred)}(\hat{m}_{11}) - H_p^{(obs)}\right)$ ($t = 600$ s). Chacune de ces valeurs est associée à une trajectoire. Différentes commandes de poussée sont testées.

masse	commande	V_a	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	*max climb*	ref	-1032	1343	1348	1693	6065
adaptative	*max climb*	ref	-758	938	1006	1205	4244
moindres carrés	*max climb*	ref	-791	966	1020	1248	7300
robuste	*max climb*	ref	-778	926	1006	1209	6742
adaptative	c_t	ref	102	928	701	933	5121
moindres carrés	c_t	ref	-117	954	696	961	6911
robuste	c_t	ref	-102	911	678	917	6360
adaptative	c_{H_p}	ref	154	928	708	940	5267
moindres carrés	c_{H_p}	ref	-119	967	705	974	7097
robuste	c_{H_p}	ref	-103	922	686	927	6518
adaptative	$c_{variante}[c_t]$	ref	113	880	666	887	4534
moindres carrés	$c_{variante}[c_t]$	ref	-116	931	677	938	6821
robuste	$c_{variante}[c_t]$	ref	-101	886	659	892	6265
adaptative	$c_{variante}[c_{H_p}]$	ref	150	887	672	899	4555
moindres carrés	$c_{variante}[c_{H_p}]$	ref	-120	940	681	947	7035
robuste	$c_{variante}[c_{H_p}]$	ref	-104	892	662	898	6458
GBM	*max climb*	ref	-350	829	710	900	4305
$mass_{ref}$	*max climb*	GBM	-715	1322	1167	1502	6050
adaptative	*max climb*	GBM	-440	841	760	949	3889
moindres carrés	*max climb*	GBM	-473	877	769	996	6432
robuste	*max climb*	GBM	-459	837	752	955	5785
adaptative	c_t	GBM	367	828	678	906	4798
moindres carrés	c_t	GBM	155	867	621	881	6048
robuste	c_t	GBM	169	824	605	841	5428
adaptative	c_{H_p}	GBM	432	832	706	937	4946
moindres carrés	c_{H_p}	GBM	165	877	627	892	6219
robuste	c_{H_p}	GBM	180	832	610	851	5569
adaptative	$c_{variante}[c_t]$	GBM	380	776	644	864	4252
moindres carrés	$c_{variante}[c_t]$	GBM	158	842	595	856	5914
robuste	$c_{variante}[c_t]$	GBM	172	797	578	815	5305
adaptative	$c_{variante}[c_{H_p}]$	GBM	431	785	670	895	4268
moindres carrés	$c_{variante}[c_{H_p}]$	GBM	168	849	598	865	6119
robuste	$c_{variante}[c_{H_p}]$	GBM	184	801	580	821	5475
GBM	*max climb*	GBM	-33.9	766	565	766	3952

masse, le modèle GBM appris dans la section 6.2 a été testé. Les tableaux 6.20 et 6.21 présentent les résultats obtenus avec les différents paramètres sur les trajectoires Mode-C et Mode-S.

Les méthodes présentées dans le chapitre 5 améliorent la prédiction. En effet, lorsque l'on compare avec les paramètres de référence, la RMSE est réduite de 49 % pour les trajectoires Mode-C et de 55 % pour les trajectoires Mode-S. Entre les méthodes, on retrouve la hiérarchie esquissée dans le chapitre 5.

6.5.3 Discussion sur les biais possibles

En regardant les résultats des deux sections précédentes, l'application directe de méthodes d'apprentissage et l'utilisation de BADA avec des paramètres prédits donne des résultats similaires avec un léger avantage pour cette dernière approche. Cette comparaison souffre toutefois de plusieurs biais. En effet, la route et l'angle d'inclinaison observés sur les futurs points sont utilisés tels quels pour calculer la trajectoire prédite. Ces informations ne sont pas retranscrites dans les variables des méthodes d'apprentissage.

En outre, les données météo ne sont pas exploitées de la même façon dans l'apprentissage et dans le calcul de trajectoire BADA. Pour les trajectoires Mode-C, seules sont incluses dans les variables w les grandeurs météo calculées avec une longitude, latitude et date identiques à celle du dernier point passé. Dans le calcul BADA, la grille est « interrogée » au fur et à mesure du calcul de la trajectoire. A priori, ce biais reste minime car les valeurs dans la grille ne varient pas beaucoup sur la distance que l'avion a le temps de parcourir en 10 min.

Dans les trajectoires Mode-S, on ne dispose pas d'une grille météo. On a uniquement les mesures à bord de l'avion. Ainsi, aucune variable dans *cmprt* ne renseigne sur la météo pour des altitudes au-dessus de $H_p\,(t=0)$. Toutefois, on se sert des points futurs pour construire une grille unidimensionnelle indexée par l'altitude pression H_p. Cette grille est utilisée dans le calcul de la trajectoire BADA. Ainsi le biais introduit dans la comparaison entre les méthodes d'apprentissage directes et celle s'appuyant sur BADA est sans doute un peu plus grand pour les trajectoires Mode-S que les trajectoires Mode-C.

Les biais précédents portent sur la différence d'informations utilisées dans les deux approches. Un troisième biais, méthodologique cette fois, existe. Le choix d'utiliser GBM parmi toutes les méthodes d'apprentissage est fait car c'est celui qui minimise l'erreur sur les 10 *ensembles de validation* de la validation croisée. Cette méthode de sélection est répétée aussi bien pour choisir la méthode prédisant la masse que celles prédisant le couple $(cas, Mach)$. En toute rigueur, si l'on souhaite utiliser les données pour choisir une méthode plutôt qu'une autre, il faut utiliser les *ensembles de test*. Ceci peut conduire à des choix de méthodes différentes pour les trois variables à prédire et pour les 10 *ensembles de validation*. Ceci n'a pas été fait pour des raisons pratiques. Ce biais est *a priori* extrêmement limité vu le faible nombre de méthodes comparées.

6.6 Conclusion

L'utilisation du modèle physique BADA requiert de connaître la masse, le profil de vitesse suivi et la commande de poussée appliquée par le pilote. Ces paramètres sont inconnus dans les systèmes au sol. Ainsi, si l'on veut calculer une prédiction avec les informations disponibles au sol, on est obligé d'utiliser les valeurs de référence contenues dans les fichiers BADA.

Ce chapitre s'intéresse à l'utilisation des informations disponibles au sol pour choisir des paramètres représentatifs de la trajectoire future de l'avion considéré.

Dans la section 6.2, plutôt que d'estimer la masse sur les points passés comme en 5.3, on cherche à apprendre sur une base d'exemples un modèle fournissant une masse dont on espère qu'elle nous donne la meilleure prévision possible de la trajectoire future. Dans les exemples utilisés pour calibrer le modèle, la réponse souhaitée du modèle est la masse ajustée sur les points futurs en utilisant la méthode décrite en 5.3. Sachant le profil de vitesse, cette méthode permet de réduire jusqu'à 66 % la RMSE sur l'altitude à $t = 600$ s.

Les masses avec lesquelles on travaille dans les sections 5 et 6.2 sont estimées en supposant une commande de poussée *max climb*. Dans la section 6.3, on utilise BFGS couplé à la méthode d'estimation des masses de la section 5.3 pour apprendre, à partir d'un jeu de trajectoires, une commande de poussée minimisant l'erreur faite sur l'excès de puissance spécifique. Différentes commandes extrêmement simples sont testées. Sachant le profil de vitesse, cette méthode permet de réduire jusqu'à 60 % la RMSE sur l'altitude à $t = 600$ s. Une notion de distance entre une commande de poussée et une trajectoire observée ressort de ces travaux. Celle-ci pourrait être exploitée pour obtenir un partitionnement du jeu de trajectoires pour lequel chaque partie contient des trajectoires avec une commande de poussée similaire.

Le profil de vitesse est étudié dans la section 6.4. Dans cette section, un algorithme détaillé dans l'annexe B détermine efficacement les valeurs $(cas, Mach)$ minimisant la RMSE entre la vitesse observée et un profil de type (cas,Mach). Cet algorithme est appliqué sur les 10 min de trajectoires à prévoir. Ceci permet de constituer un ensemble d'apprentissage qui va permettre d'apprendre des modèles reliant les variables connues aux valeurs cas et $Mach$. Des méthodes d'apprentissage sont appliquées sur chacune de ces variables. L'utilisation conjointe de ces deux modèles appris séparément permet de réduire la RMSE sur la vitesse de 35 %.

Dans la section 6.5 on a appliqué des méthodes d'apprentissage pour directement apprendre l'altitude finale $H_p\,(t = 600 \text{ s})$. Nous avons ensuite comparé les différentes approches : apprentissage direct de l'altitude, apprentissage des paramètres du modèle physique, méthode de référence BADA avec la masse de référence ou les masses estimées par les méthodes estimées présentées au chapitre 5. Les deux approches réduisent considérablement l'erreur sur l'altitude par rapport à BADA utilisé avec les paramètres de référence : la RMSE sur l'altitude finale est réduite d'environ 50 %. Entre ces deux approches, un léger avantage apparaît pour le modèle physique utilisé avec les paramètres appris.

Chapitre 7

Généralisation à différents types avions et différentes altitudes initiales

Sommaire

7.1 Des altitudes initiales différentes 169
7.1.1 Construction du jeu d'exemples 169
7.2 Résultats 171
7.2.1 Étude sur les exemples construits 171
7.2.2 Étude sur des segments de montée plus petits 181
7.3 Conclusion 184

Dans tout ce qui précède, on a considéré un avion de type A320 à une altitude initiale $H_p\,(t=0) = 18\ 000$ ft. Ce cas ne recouvre pas toutes les trajectoires en montée dans l'espace aérien supérieur. On a donc élargi notre étude à différents types d'avions et différentes altitudes initiales $H_p\,(t=0)$.

7.1 Des altitudes initiales différentes

Le filtrage des trajectoires en phase de montée est identique à celui présenté dans la sous-section 4.1.4. La construction des exemples est toutefois différente. On utilise les trajectoires Mode-C d'avions en partance des deux grands aéroports parisiens LFPG et LFPO.

7.1.1 Construction du jeu d'exemples

Dans la sous-section 4.1.5, la fonction $H_p\,(t)$ permet de déterminer la date t_0 telle que $H_p\,(t_0) = 18\ 000$ ft. La trajectoire est échantillonnée par rapport à cette date t_0 en prenant $p = 11$ points passés et $q = 40$ points futurs. Ainsi chaque trajectoire donne lieu à la création d'un seul et unique exemple. Un exemple étant constitué de $p + q$ points consécutifs d'une trajectoire en phase de montée.

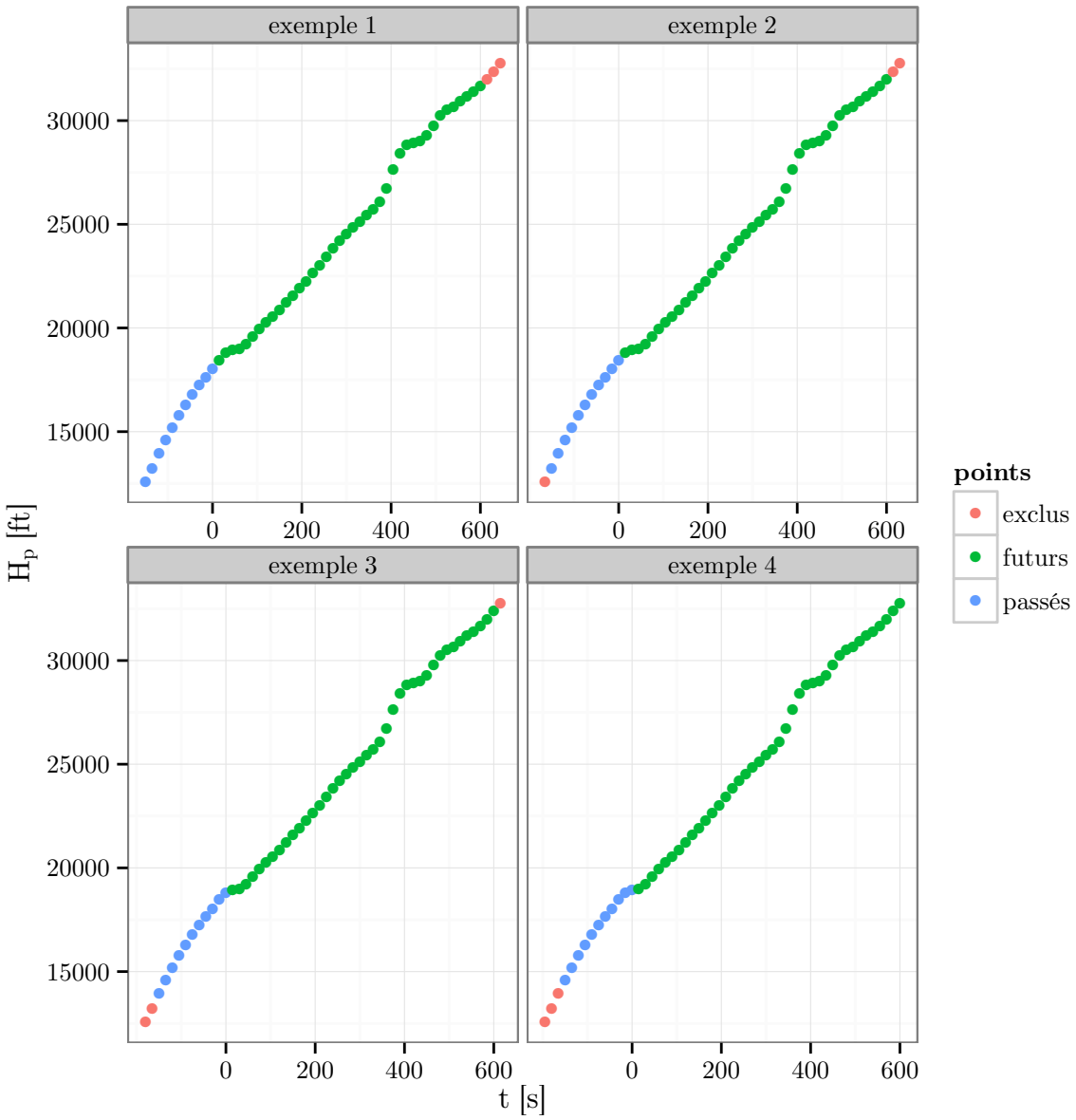


FIGURE 7.1 – À partir d'une trajectoire de 54 points, on construit 4 exemples de 51 points chacun. L'origine du temps t est décalée d'un point pour générer chaque exemple.

Ici, les trajectoires ne sont plus échantillonnées à partir de la date de passage du FL180 t_0. Ils sont échantillonnés à partir d'une date fixe quelconque. On obtient ainsi une séquence de n points espacés de 15 s. Si $n \geqslant p + q$, on peut extraire $n - p - q + 1$ exemples d'une trajectoire de n points. Suivant la plage d'altitude initiale H_{p_0} qui nous intéresse, on peut garder uniquement les exemples dont le $p^{ième}$ point est supérieur à une altitude donnée. Dans notre cas, on a gardé les exemples intervenant dans l'espace aérien supérieur, soit une altitude H_{p_0} supérieure à 18 000 ft. Pour certains types avions, il y a peu d'avions qui montent encore pendant 10 min après avoir passé 18 000 ft. Pour contourner ce problème, on a réduit le seuil de 18 000 ft à 15 000 ft pour les types avions concernés. La figure 7.1 illustre le mécanisme de « fenêtre glissante » qui permet de construire des exemples ayant une altitude initiale différente à partir d'une trajectoire.

Les exemples construits précédemment forment notre base d'exemples pour l'apprentissage. Avec ce mécanisme de « fenêtre glissante », plusieurs exemples peuvent provenir de la même montée d'un même avion. Deux exemples « consécutifs » partagent 50 points d'une même trajectoire. Ceci peut être gênant dans le contexte de l'apprentissage. Pour remédier à ce problème, chaque partition d'un jeu d'exemples est faite en se basant sur les trajectoires. Ainsi, dans une partition, tout les exemples d'une même trajectoire sont dans la même partie. Ceci permet d'éviter que des exemples dans l'*ensemble d'apprentissage* et l'*ensemble de validation* proviennent de trajectoires identiques.

7.2 Résultats

Ces jeux d'exemples sont exploités avec GBM et les variables $c\hat{m}prw$ pour apprendre la masse future $\hat{m}_{11,futur}$ et le couple $(cas, Mach)$. La méthode employée est identique à celle décrite dans les sections 6.2 et 6.4. Les méthodes, décrites dans la section 5, permettant d'estimer la masse sont également utilisées sur les exemples. La méthode permettant d'apprendre une commande de poussée décrite dans la section 6.3 est aussi testée. Elle n'est toutefois pas testée sur les exemples mais directement sur les trajectoires. Ceci permet de limiter le nombre de trajectoires prises en compte pour l'apprentissage. Le temps de calcul lors de la phase d'apprentissage s'en trouve réduit. La commande de poussée est modélisée comme fonction de l'altitude H_p. On note $BADA_{ref}$ le prédicteur utilisant BADA avec la masse et le profil de vitesse de référence renseignés dans les fichiers BADA. De même, on note $BADA_{\text{GBM}}$ le prédicteur utilisant BADA avec la masse et les valeurs $(cas, Mach)$ prédites par GBM.

7.2.1 Étude sur les exemples construits

Les tableaux 7.1 à 7.9 présentent les statistiques obtenues sur l'altitude prédite à $t =$ 600 s pour différents types avions.

Pour chacun de ces types avions, la RMSE est au moins divisée par deux lorsque l'on compare $BADA_{ref}$ et $BADA_{GBM}$.

Au regard des prédictions de $BADA_{ref}$, plus le type avion a une masse de référence

TABLE 7.1 – Statistiques sur l'erreur $\left(H_p^{(pred)} - H_p^{(obs)}\right)(t = 600 \text{ s})$ faite en prédisant à partir d'altitudes initiales $H_p(t = 0)$ variable. Les exemples sur lesquels les statistiques sont calculées sont des A319 ($mass_{ref} = 60\,000$ kg).

masse	commande	V_a	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	*max climb*	ref	274	1472	1176	1497	5315
adaptative	*max climb*	GBM	-171	938	741	953	5475
moindres carrés	*max climb*	GBM	-477	1005	834	1112	10978
robuste	*max climb*	GBM	-459	974	815	1077	10901
adaptative	c_{H_p}	GBM	771	928	936	1206	5652
moindres carrés	c_{H_p}	GBM	120	957	684	964	10881
robuste	c_{H_p}	GBM	134	927	667	937	10771
GBM	*max climb*	GBM	42.1	725	532	726	5529
prévision directe de H_p par GBM			44.3	799	596	800	5280

TABLE 7.2 – Statistiques sur l'erreur $\left(H_p^{(pred)} - H_p^{(obs)}\right)(t = 600 \text{ s})$ faite en prédisant à partir d'altitudes initiales $H_p(t = 0)$ variable. Les exemples sur lesquels les statistiques sont calculées sont des A320 ($mass_{ref} = 64\,000$ kg).

masse	commande	V_a	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	*max climb*	ref	290	1420	1165	1449	5753
adaptative	*max climb*	GBM	-156	857	671	871	6905
moindres carrés	*max climb*	GBM	-350	815	678	887	7596
robuste	*max climb*	GBM	-343	801	671	872	7915
adaptative	c_{H_p}	GBM	595	853	819	1040	7810
moindres carrés	c_{H_p}	GBM	78.3	818	607	822	7528
robuste	c_{H_p}	GBM	83.4	804	599	808	7851
GBM	*max climb*	GBM	23.5	681	490	681	7193
prévision directe de H_p par GBM			27.8	736	545	737	7328

TABLE 7.3 – Statistiques sur l'erreur $\left(H_p^{(pred)} - H_p^{(obs)}\right)(t = 600 \text{ s})$ faite en prédisant à partir d'altitudes initiales $H_p(t = 0)$ variable. Les exemples sur lesquels les statistiques sont calculées sont des A321 ($mass_{ref} = 72\,000$ kg).

masse	commande	V_a	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	*max climb*	*ref*	863	1683	1588	1891	6154
adaptative	*max climb*	GBM	-273	950	762	989	5326
moindres carrés	*max climb*	GBM	-654	857	870	1079	6082
robuste	*max climb*	GBM	-655	850	866	1073	6080
adaptative	c_{H_p}	GBM	492	903	794	1028	5354
moindres carrés	c_{H_p}	GBM	116	843	613	851	6750
robuste	c_{H_p}	GBM	113	835	605	842	6740
GBM	*max climb*	GBM	22.1	774	554	774	4418
prévision directe de H_p par GBM			27.4	849	631	849	4304

TABLE 7.4 – Statistiques sur l'erreur $\left(H_p^{(pred)} - H_p^{(obs)}\right)(t = 600 \text{ s})$ faite en prédisant à partir d'altitudes initiales $H_p(t = 0)$ variable. Les exemples sur lesquels les statistiques sont calculées sont des A332 ($mass_{ref} = 190\,000$ kg).

masse	commande	V_a	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	*max climb*	*ref*	2622	1820	2783	3192	6769
adaptative	*max climb*	GBM	919	1162	1247	1481	5620
moindres carrés	*max climb*	GBM	162	880	676	895	5566
robuste	*max climb*	GBM	162	870	668	885	5574
adaptative	c_{H_p}	GBM	781	1034	1104	1295	5522
moindres carrés	c_{H_p}	GBM	98.6	878	649	884	5886
robuste	c_{H_p}	GBM	98.8	867	641	873	5891
GBM	*max climb*	GBM	70.4	651	460	654	4934
prévision directe de H_p par GBM			34.2	696	505	697	5172

TABLE 7.5 – Statistiques sur l'erreur $\left(H_p^{(pred)} - H_p^{(obs)}\right)(t = 600 \text{ s})$ faite en prédisant à partir d'altitudes initiales $H_p(t = 0)$ variable. Les exemples sur lesquels les statistiques sont calculées sont des B737 ($mass_{ref} = 60\ 000$ kg).

masse	commande	V_a	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	*max climb*	ref	606	1750	1619	1852	4157
adaptative	*max climb*	GBM	-625	1487	1166	1613	6526
moindres carrés	*max climb*	GBM	-1496	1953	1702	2460	10812
robuste	*max climb*	GBM	-1407	1904	1626	2367	12189
adaptative	c_{H_p}	GBM	1029	1462	1537	1788	6193
moindres carrés	c_{H_p}	GBM	163	1954	1542	1960	9303
robuste	c_{H_p}	GBM	259	1902	1506	1919	11160
GBM	*max climb*	GBM	-52	804	629	805	3645
prévision directe de H_p par GBM			42.4	883	675	884	4184

TABLE 7.6 – Statistiques sur l'erreur $\left(H_p^{(pred)} - H_p^{(obs)}\right)(t = 600 \text{ s})$ faite en prédisant à partir d'altitudes initiales $H_p(t = 0)$ variable. Les exemples sur lesquels les statistiques sont calculées sont des B744 ($mass_{ref} = 285\ 700$ kg).

masse	commande	V_a	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	*max climb*	ref	5558	1646	5580	5797	10183
adaptative	*max climb*	GBM	2134	1516	2241	2618	6921
moindres carrés	*max climb*	GBM	-511	1010	841	1132	6134
robuste	*max climb*	GBM	-497	978	823	1097	5209
adaptative	c_{H_p}	GBM	2985	1471	3004	3327	7629
moindres carrés	c_{H_p}	GBM	343	970	786	1028	4714
robuste	c_{H_p}	GBM	355	935	768	1000	4281
GBM	*max climb*	GBM	142	778	586	790	3342
prévision directe de H_p par GBM			84.5	827	624	831	4947

TABLE 7.7 – Statistiques sur l'erreur $\left(H_p^{(pred)} - H_p^{(obs)}\right)(t = 600 \text{ s})$ faite en prédisant à partir d'altitudes initiales $H_p(t = 0)$ variable. Les exemples sur lesquels les statistiques sont calculées sont des B772 ($mass_{ref} = 208\ 700$ kg).

masse	commande	V_a	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	*max climb*	ref	3728	1413	3750	3987	7145
adaptative	*max climb*	GBM	919	993	1092	1353	3865
moindres carrés	*max climb*	GBM	-112	685	523	694	5509
robuste	*max climb*	GBM	-113	675	519	684	5526
adaptative	c_{H_p}	GBM	1310	969	1366	1630	4305
moindres carrés	c_{H_p}	GBM	135	671	497	684	5496
robuste	c_{H_p}	GBM	134	660	491	673	5508
GBM	*max climb*	GBM	112	500	385	512	3446
prévision directe de H_p par GBM			36.8	543	416	544	2676

TABLE 7.8 – Statistiques sur l'erreur $\left(H_p^{(pred)} - H_p^{(obs)}\right)(t = 600 \text{ s})$ faite en prédisant à partir d'altitudes initiales $H_p(t = 0)$ variable. Les exemples sur lesquels les statistiques sont calculées sont des E145 ($mass_{ref} = 18\ 500$ kg).

masse	commande	V_a	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	*max climb*	ref	1623	1801	1909	2425	7428
adaptative	*max climb*	GBM	-545	1526	1278	1620	6721
moindres carrés	*max climb*	GBM	-1023	1554	1502	1860	7622
robuste	*max climb*	GBM	-1024	1553	1502	1860	7438
adaptative	c_{H_p}	GBM	789	1453	1283	1653	7331
moindres carrés	c_{H_p}	GBM	126	1503	1140	1508	7905
robuste	c_{H_p}	GBM	126	1499	1137	1505	7742
GBM	*max climb*	GBM	190	1314	1010	1327	5378
prévision directe de H_p par GBM			45.4	1283	979	1284	5493

TABLE 7.9 – Statistiques sur l'erreur $\left(H_p^{(pred)} - H_p^{(obs)}\right)(t = 600 \text{ s})$ faite en prédisant à partir d'altitudes initiales $H_p(t = 0)$ variable. Les exemples sur lesquels les statistiques sont calculées sont des F100 ($mass_{ref} = 38\ 000$ kg).

masse	commande	V_a	moy	écart-type	moy abs	rmse	max abs
$mass_{ref}$	*max climb*	ref	556	1879	1616	1959	6539
adaptative	*max climb*	GBM	-41.6	1278	1037	1278	4453
moindres carrés	*max climb*	GBM	-160	1206	971	1217	5424
robuste	*max climb*	GBM	-159	1198	967	1208	4466
adaptative	c_{H_p}	GBM	867	1228	1225	1503	5787
moindres carrés	c_{H_p}	GBM	157	1156	925	1167	5255
robuste	c_{H_p}	GBM	157	1146	921	1157	4383
GBM	*max climb*	GBM	102	1022	793	1027	4490
prévision directe de H_p par GBM			39.3	1102	862	1103	4454

élevée, plus celle-ci semble sous-estimer la vraie masse de l'avion lors de la phase de montée initiale. Ceci se confirme en comparant les valeurs de masses estimées aux valeurs de références. Par exemple, pour les avions de type B744 la masse de référence m_{ref} vaut 285 700 kg alors que la moyenne des masses $\hat{m}_{11,futur}$ est 372 200 kg. Ceci n'est pas étonnant car on observe un avion en montée venant juste de décoller et devant parcourir en moyenne 4 749 Nm.

Cette observation explique les faibles performances de la méthode adaptative sur les avions avec une grande masse de référence. En effet, cette dernière utilise la masse de référence pour première estimation. Ensuite, cette estimation va être mise à jour mais l'incrément reste limité à 2 % de la masse de référence. De plus, la masse estimée est bornée à plus ou moins 20 % de la masse de référence. Dans le cas du B744, on a une différence de 30 % entre la masse de référence et la moyenne des masses $\hat{m}_{11,futur}$. Ainsi, la méthode adaptative estimera une masse toujours inférieure à la moyenne des masses $\hat{m}_{11,futur}$.

De façon générale, les résultats obtenus avec les méthodes d'estimation de la masse sont contrastés entre les différents types avion. Pour le B772, les performances des méthodes d'estimation sont assez proches de celles obtenues avec la prédiction de la masse par GBM. Au contraire, pour le B737, les méthodes d'estimation de la masse donnent une estimation conduisant à des performances inférieures à celle de la masse de référence. Ceci est d'autant plus étonnant que GBM fournit des prédictions satisfaisantes. Toutefois, dans l'ensemble, sur les autres types avions, ces méthodes améliorent les résultats comparés à $BADA_{ref}$.

Concernant les commandes de poussée apprises, elles améliorent toujours le résultat des méthodes moindres carrés et robuste. À l'inverse, les résultats sont toujours dégradés pour la méthode adaptative.

La méthode consistant à prédire directement l'altitude $H_p(t = 600 \text{ s})$ avec GBM et les variables $\hat{cmprw}$ est aussi évaluée dans la dernière des tableaux 7.1 à 7.9. La prédiction faite avec ce modèle est meilleure que celle des méthodes BADA avec une estimation de la

masse. Toutefois, les résultats sont légèrement moins bon que ceux de $BADA_{\mathrm{GBM}}$.

La figure 7.2 représente les fonctions de répartition de la valeur absolue de l'erreur obtenue avec $BADA_{ref}$ ou $BADA_{\mathrm{GBM}}$ pour différents types d'avion. Avec cette figure on confirme que $BADA_{\mathrm{GBM}}$ est plus performant que $BADA_{ref}$ et ce pour tout quantile d'erreur. L'écart entre les deux méthodes est le plus grand lorsque l'on s'intéresse à des quantiles proches des 90 %.

La figure 7.3 présente la valeur absolue de l'erreur en fonction de l'altitude initiale de la prédiction. L'erreur semble décroître avec l'altitude $H_p\,(t=0)$. Il faudrait toutefois faire des tests pour vérifier plus formellement cette affirmation. Toutefois cette affirmation est supportée par la physique du problème. L'excès de puissance spécifique d'un avion donné diminue avec l'altitude. Ainsi, à vitesse constante, le taux de montée maximal diminue avec l'altitude. Cette diminution est observée même dans la plage $Mach$ constante dans laquelle de la vitesse est échangée contre un plus grand taux de montée. La plage dans laquelle peut évoluer le taux de montée d'un avion diminue avec l'altitude. Ainsi, intuitivement, on peut penser que ceci participe à une réduction de l'erreur pour la prédiction de trajectoire pour des altitudes élevées. Cette figure permet également de voir la distribution des altitudes initiales dans nos exemples.

La figure 7.4 présente l'erreur obtenue en fonction du temps avec $BADA_{ref}$ ou $BADA_{\mathrm{GBM}}$ pour différents types avions. Toutes les 15 s, la distribution de l'erreur $\left(H_p^{(pred)} - H_p^{(obs)}\right)$ est résumée par une boîte à moustaches qui permet de visualiser de bas en haut les quantiles 5 %, 25 %, 50 %, 75 % et 95 %. Visuellement, pour chaque méthode, deux cônes imbriqués apparaissent, le cône le plus intérieur contient 50 % des erreurs tandis que le cône extérieur contient 90 % des erreurs. Avec cette figure on retrouve que la masse des avions A332, B772 et B744 est sous-estimée par la masse de référence $mass_{ref}$. Pour la méthode $BADA_{\mathrm{GBM}}$, la médiane de l'erreur reste centrée sur toute la prévision. Le gain apporté par $BADA_{\mathrm{GBM}}$ pour les avions A319, A320 et A321 est faible sur les deux premières minutes.

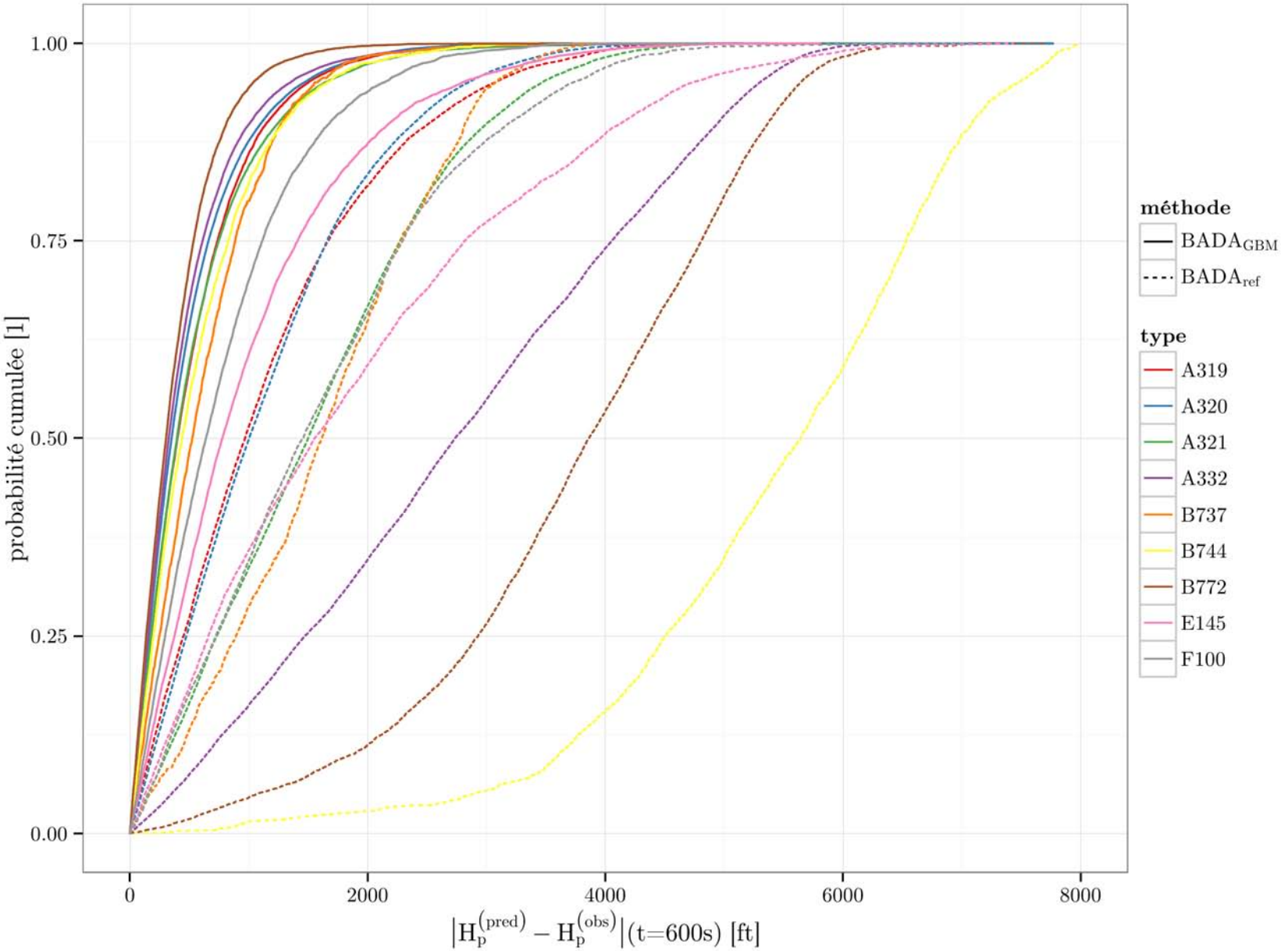


FIGURE 7.2 – Cette figure présente les fonctions de répartitions pour la valeur absolue de l'erreur obtenue avec $BADA_{ref}$ ou $BADA_{\mathrm{GBM}}$ pour différents types avions.

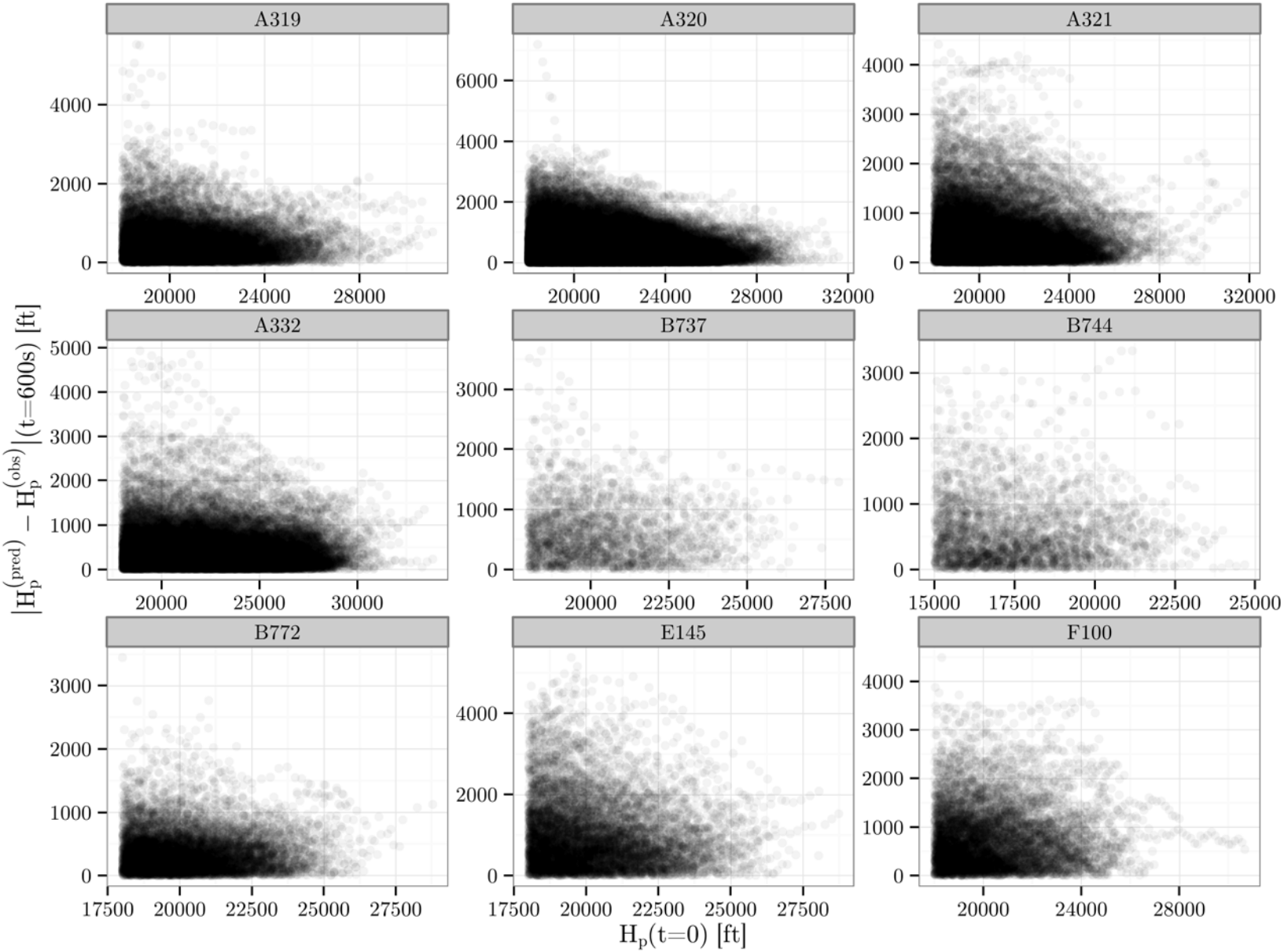


FIGURE 7.3 – Cette figure présente la valeur absolue de l'erreur obtenue avec $BADA_{\mathrm{GBM}}$ en fonction de l'altitude initiale de la prédiction.

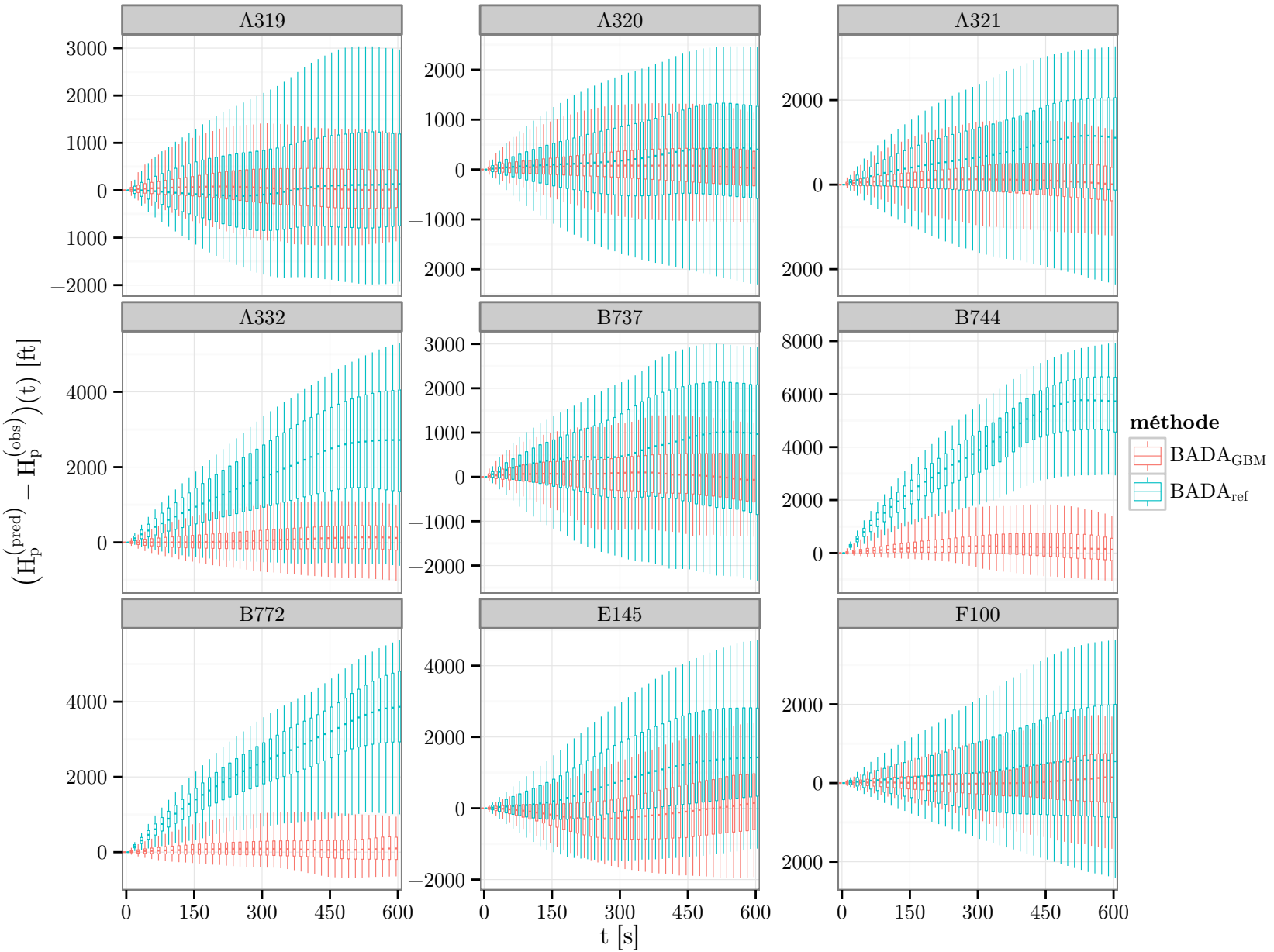


FIGURE 7.4 – Cette figure présente l'erreur obtenue en fonction du temps avec $BADA_{ref}$ ou $BADA_{GBM}$ pour différents types avions. Toutes les 15 s, la distribution de l'erreur $\left(H_p^{(pred)} - H_p^{(obs)}\right)$ est résumée par une boîte à moustaches qui permet de visualiser de bas en haut les quantiles 5 %, 25 %, 50 %, 75 % et 95 %.

7.2.2 Étude sur des segments de montée plus petits

Jusqu'à présent on s'est intéressé à la prédiction de montée avec un horizon de 10 min. Ainsi, seuls les segments de montée faisant au moins 10 min ont été utilisés dans la construction des exemples. Le tableau 7.10 regroupe les différentes statistiques sur le nombre de segments de montée. Avec ce tableau, on remarque qu'une grande partie des segments de montée ont été écartés de cette étude. On note *moins_de_10min* l'ensemble des segments avec un futur d'au moins 5 min mais inférieur à 10 min et *plus_de_10min* l'ensemble des segments avec un futur d'au moins 10 min. Ces deux ensembles forment une partition des segments avec un futur d'au moins 5 min. Les prédicteurs s'appuyant sur BADA mais entraînés sur des montées de 10 min peuvent prédire des trajectoires pour des horizons plus faibles. Si le segment de montée contient onze points passés alors la prédiction peut être calculée mais ce n'est pas parce qu'elle calculable qu'elle est correcte. Dans cette sous-section, en utilisant ces prédicteurs, on va s'intéresser à la qualité de la prédiction sur un horizon de 5 min de trajectoires issues de *moins_de_10min* et de *plus_de_10min*. On va ainsi pouvoir vérifier si un prédicteur entraîné sur des montées de 10 min est efficace sur une part importante des segments de montée observés dans l'espace aérien supérieur.

TABLE 7.10 – Statistiques sur le nombre de segments de montée.

type	seuil [ft]	nombre de segments de montée supérieurs au seuil	nombre de segments de montée supérieurs au seuil avec un passé suffisant	nombre de segments de montée supérieurs au seuil avec un passé suffisant et un futur d'au moins 10 min	nombre de segments de montée supérieurs au seuil avec un passé suffisant et un futur d'au moins 5 min et d'au plus 10 min
A319	18000	9250	8423	1863	3349
A320	18000	16939	15486	5729	5621
A321	18000	4050	3789	1866	967
A332	18000	2459	2166	1475	326
B737	18000	1752	1512	344	481
B744	15000	2083	1673	350	888
B772	18000	2565	2146	910	551
E145	18000	3249	2779	851	1097
F100	18000	2756	2565	660	1070

On note x les variables explicatives i.e. la trajectoire passée, y les 5 premières minutes de la trajectoire future, $p_{moins}(x, y)$ et $p_{plus}(x, y)$ les probabilités sur les exemples issus respectivement de *moins_de_10min* et *plus_de_10min*. L'ensemble *plus_de_10min* regroupe les segments ayant servi à construire les exemples utilisés dans l'apprentissage. Les

probabilités $p_{moins}(x, y)$ et $p_{plus}(x, y)$ peuvent être très différentes. Par conséquent, bien prédire sur *plus_de_10min* ne donne aucune garantie sur la qualité de la prédiction pour des segments de *moins_de_10min*. De plus, lors de l'apprentissage, tous les exemples dans l'*ensemble d'apprentissage* proviennent de *plus_de_10min*, la probabilité $p_{moins}(x, y)$ n'a donc aucun impact sur le résultat de l'apprentissage.

Les probabilités $p_{moins}(x, y)$ et $p_{plus}(x, y)$ peuvent différer de diverses manières. Par exemple, pour un x donné, rien ne dit *a priori* que les pilotes opèrent les 5 premières minutes de montée de la même façon suivant que la montée complète dure 5 min ou 10 min. Énoncé plus formellement, on peut avoir $p_{moins}(y|x) \neq p_{plus}(y|x)$. Toutefois, même si $p_{moins}(y|x) = p_{plus}(y|x)$, on peut avoir $p_{moins}(x) \neq p_{plus}(x)$. Dans l'apprentissage, lorsque un tel décalage arrive entre l'*ensemble d'apprentissage* et l'*ensemble de validation*, on parle de *covariate shift*. Ce décalage oblige à corriger le risque empirique avec le ratio $p_{validation}(x)/p_{apprentissage}(x)$ ([Sugiyama 07]). Dans notre cas ceci conduirait à des statistiques différentes sur l'erreur pour *moins_de_10min* et *plus_de_10min*, et ce même si la relation liant les y et x était identique. Ainsi, il sera difficile de conclure sur la similarité des trajectoires en regardant uniquement le tableaux d'erreurs 7.11. Ceci rappelle que spécifier le concept que l'on souhaite apprendre ne se résume pas à simplement spécifier un y et un x mais aussi et surtout à sélectionner un ensemble d'exemples représentatifs de l'application envisagée. Dans notre cas, l'application envisagée est la prédiction de 10 min de montée.

Même si le prédicteur $BADA_{\text{GBM}}$ est « spécialisé » pour la prédiction de 10 min de montée, le tableau d'erreurs 7.11 permet d'évaluer en pratique l'erreur obtenue en utilisant $BADA_{\text{GBM}}$ pour prédire 5 min de montée sur les segments *moins_de_10min*. Avec ce tableau, on remarque que les statistiques sont très différentes entre les segments *moins_de_10min* et *plus_de_10min*. Sur les segments *moins_de_10min* et *plus_de_10min*, l'utilisation du prédicteur $BADA_{\text{GBM}}$ réduit toujours l'erreur faite comparée à celle faite avec $BADA_{ref}$. Ce constat confirme l'intérêt de l'utilisation du prédicteur $BADA_{\text{GBM}}$ pour prédire sur des segments de moins de 10 min.

Dans tout ce qui précède, on a supposé que l'on dispose d'un passé suffisant ce qui correspond à avoir un taux de montée supérieur à 150 ft/min sur les 205 secondes précédant $t = 0$. Ce critère sert à identifier un passé correspondant à une phase de montée. Si le passé n'est pas une phase de montée, le prédicteur $BADA_{\text{GBM}}$ peut tout de même calculer une prédiction mais celle-ci ne sera vraisemblablement pas correcte. Par exemple, un avion en palier volant à vitesse constante a une variation d'énergie spécifique nulle. Or, dans l'*ensemble d'apprentissage*, de faibles valeurs de variation d'énergie spécifique sont associées à une masse $\hat{m}_{11,futur}$ élevée. Ainsi, il est vraisemblable que la masse prédite sur des trajectoires en palier surestime la masse $\hat{m}_{11,futur}$. Pour prédire la trajectoire de montée d'un avion en palier à $t = 0$, on pourrait remonter plus loin dans le passé pour identifier un segment de montée à partir duquel on prédirait $(\hat{m}_{11,futur}, cas, Mach)$. Ensuite, avec le modèle de consommation carburant, on pourrait mettre à jour l'estimation de la masse pour qu'elle prenne en compte le carburant consommé depuis le segment de montée jusqu'à $t = 0$. Tout ceci doit être confirmé par une étude plus détaillée.

TABLE 7.11 – Statistiques sur l'erreur $\left(H_p^{(pred)} - H_p^{(obs)}\right)$ ($t = 300$ s) sur des segments de montée de longueur inférieure et supérieure à 10 min.

type	données	prédicteur	moy	écart-type	moy abs	rmse	max abs
A319	*plus_de_10min*	$BADA_{ref}$	9.77	1184	966	1184	4691
A319	*plus_de_10min*	$BADA_{\mathrm{GBM}}$	148	754	534	768	4215
A319	*moins_de_10min*	$BADA_{ref}$	-572	1294	1165	1415	5769
A319	*moins_de_10min*	$BADA_{\mathrm{GBM}}$	6.14	942	662	941	5338
A320	*plus_de_10min*	$BADA_{ref}$	152	1073	855	1084	5557
A320	*plus_de_10min*	$BADA_{\mathrm{GBM}}$	164	686	463	705	5595
A320	*moins_de_10min*	$BADA_{ref}$	-446	1286	1090	1361	5378
A320	*moins_de_10min*	$BADA_{\mathrm{GBM}}$	24.5	921	630	921	4382
A321	*plus_de_10min*	$BADA_{ref}$	555	1201	1082	1323	4941
A321	*plus_de_10min*	$BADA_{\mathrm{GBM}}$	214	734	515	765	4461
A321	*moins_de_10min*	$BADA_{ref}$	15.8	1389	1152	1388	4122
A321	*moins_de_10min*	$BADA_{\mathrm{GBM}}$	13.3	894	615	893	4075
A332	*plus_de_10min*	$BADA_{ref}$	1664	1182	1796	2042	4694
A332	*plus_de_10min*	$BADA_{\mathrm{GBM}}$	99.6	549	386	558	3728
A332	*moins_de_10min*	$BADA_{ref}$	962	1437	1499	1727	3728
A332	*moins_de_10min*	$BADA_{\mathrm{GBM}}$	-25.4	641	448	641	2822
B737	*plus_de_10min*	$BADA_{ref}$	399	1261	1101	1322	3780
B737	*plus_de_10min*	$BADA_{\mathrm{GBM}}$	124	727	538	737	3697
B737	*moins_de_10min*	$BADA_{ref}$	-409	1119	956	1190	3732
B737	*moins_de_10min*	$BADA_{\mathrm{GBM}}$	-157	776	595	791	3908
B744	*plus_de_10min*	$BADA_{ref}$	3807	1169	3819	3982	6530
B744	*plus_de_10min*	$BADA_{\mathrm{GBM}}$	442	677	577	808	3488
B744	*moins_de_10min*	$BADA_{ref}$	2317	1296	2461	2655	6133
B744	*moins_de_10min*	$BADA_{\mathrm{GBM}}$	-95.7	847	601	852	4555
B772	*plus_de_10min*	$BADA_{ref}$	2305	911	2332	2479	4925
B772	*plus_de_10min*	$BADA_{\mathrm{GBM}}$	183	448	324	484	2874
B772	*moins_de_10min*	$BADA_{ref}$	1750	1193	1918	2117	4103
B772	*moins_de_10min*	$BADA_{\mathrm{GBM}}$	89	577	388	584	3555
E145	*plus_de_10min*	$BADA_{ref}$	828	1388	1308	1616	4964
E145	*plus_de_10min*	$BADA_{\mathrm{GBM}}$	-169	916	724	932	4228
E145	*moins_de_10min*	$BADA_{ref}$	-414	1613	1330	1665	5076
E145	*moins_de_10min*	$BADA_{\mathrm{GBM}}$	-763	1092	1084	1332	3956
F100	*plus_de_10min*	$BADA_{ref}$	244	1125	961	1151	3604
F100	*plus_de_10min*	$BADA_{\mathrm{GBM}}$	57.2	604	414	607	3453
F100	*moins_de_10min*	$BADA_{ref}$	-150	1095	876	1105	3447
F100	*moins_de_10min*	$BADA_{\mathrm{GBM}}$	-184	644	453	670	4267

7.3 Conclusion

Jusqu'à présent l'étude portait sur des segments de montée de plus de 10 min avec des avions de type A320 et une altitude initiale $H_p\,(t=0) = 18\ 000$ ft. Dans ce chapitre, on élargit cette étude pour savoir si nos méthodes sont capables d'améliorer la prédiction de n'importe quel segment de montée situé dans l'espace aérien supérieur.

Dans un premier temps, nous avons appliqué les méthodes développées dans les chapitres précédents à différents types avions et à différentes altitudes initiales $H_p\,(t=0)$ sur des segments de montée de plus de 10 min situés dans l'espace aérien supérieur. Les résultats confirment l'amélioration apportée par la prédiction de la masse et du couple $(cas, Mach)$, comparée à l'utilisation des valeurs de référence dans les fichier BADA. Dans un second temps, après avoir constaté que tous les segments de montée ne durent pas 10 min, nous avons étudié la prédiction sur un horizon de 5 min à l'aide d'un prédicteur entraîné à prédire à un horizon de 10 min sur des segments de plus de 10 min. On a distingué les segments de plus de 10 min des segments de plus de 5 min mais de moins de 10 min. Les prédictions à 5 min sont calculées pour chacun de ces deux ensembles. Sur ces deux ensembles, on note une amélioration avec la prédiction de la masse et du couple $(cas, Mach)$. Toutefois, les erreurs sont plus grandes sur les segments de moins de 10 min que ceux de plus de 10 min. Il est difficile d'expliquer cet écart qui peut avoir plusieurs causes. Il peut provenir du fait que les avions avec un même passé opèrent différemment suivant la durée de montée prévue (*i.e.* $p_{moins}(y|x) \neq p_{plus}(y|x)$). Cet écart peut également provenir du fait que le passé des trajectoires est distribué différemment dans les deux cas (*i.e.* $p_{moins}(x) \neq p_{plus}(x)$).

Ces différents cas étudiés confirment l'intérêt des méthodes développées pour améliorer la prédiction des segments de montée dans l'espace aérien supérieur.

Toutefois, tous les cas n'ont pas été étudiés. Il reste à étudier les cas où le passé n'est pas une phase de montée et les cas où un palier est inclus dans l'horizon de prédiction.

Conclusion générale

Nous avons proposé dans cette thèse de nouvelles méthodes pour la prédiction, à court-moyen terme, de la trajectoire d'un avion en montée avec les informations actuellement disponibles au sol. Notre étude s'est focalisée sur l'espace aérien supérieur. Plutôt que d'appliquer des méthodes statistiques directement sur les trajectoires, en oubliant complètement toute modélisation physique, nous avons choisi d'utiliser des méthodes statistiques pour inférer les paramètres non disponibles au sol et nécessaires à l'utilisation d'un modèle physique. Parmi ces paramètres, nous avons cherché à acquérir des informations sur la masse, la commande de poussée et le profil de vitesse.

Contributions

Concernant l'estimation de la masse, une nouvelle approche reposant entièrement sur le modèle physique est développée. La masse estimée est celle minimisant l'écart entre la variation d'énergie spécifique observée et l'excès de puissance spécifique calculé. Cette approche est appliquée sur les points passés de la trajectoire ce qui permet d'obtenir une estimation de la masse à partir de la trajectoire passée. Cette méthode est comparée sur des trajectoires simulées et réelles à la méthode adaptative développé par [Schultz 12]. Les performances obtenues sont du même ordre de grandeur pour les deux méthodes avec un avantage pour la notre sur les trajectoires simulées et un léger avantage pour la méthode adaptative sur les trajectoires réelles, sauf lorsque celle-ci est combinée avec l'apprentissage du profil de poussée.

Cette nouvelle méthode est également appliquée sur les points futurs des trajectoires. Ces masses estimées sur les points futurs permettent de construire un jeu d'exemples utilisé par nos méthodes d'apprentissage. Cette base d'exemples permet d'apprendre des modèles reliant la masse ajustée sur les points futurs aux variables disponibles lors du calcul de la prédiction. Contrairement aux méthodes évoquées précédemment qui utilisent uniquement le modèle physique, un jeu d'exemples est ici utilisé ainsi que des variables n'apparaissant pas dans le modèle physique. Cette méthode permet d'obtenir de meilleures prédictions que toutes les méthodes testées dans cette thèse.

Les masses sont estimées en supposant une commande de poussée. En utilisant la méthode permettant d'estimer la masse minimisant l'écart entre la variation d'énergie spécifique observée et l'excès de puissance spécifique calculé, on peut apprendre une commande de poussée minimisant l'erreur faite sur l'excès de puissance spécifique. Le gain associé à cette

méthode est toutefois variable suivant la fonction paramétrique choisie pour la commande. Différentes commandes extrêmement simples sont testées avec des résultats variables. Dans le meilleur cas, en exploitant la variante de l'appareil, on arrive à des résultats presque aussi bon que ceux obtenus en prédisant la masse estimée sur les points futurs.

Pour le profil de vitesse, un algorithme efficace détermine les valeurs $(cas, Mach)$ minimisant la RMSE entre la vitesse observée et un profil de type (cas,Mach). Cet algorithme développé dans cette thèse est appliqué sur les points futurs des trajectoires. Ceci permet de constituer un *ensemble d'apprentissage* qui va permettre d'apprendre des modèles reliant les variables connues aux valeurs cas et $Mach$. Des méthodes d'apprentissage sont appliquées sur chacune de ces variables.

Si l'on combine toutes ces méthodes pour choisir les paramètres, suivant le type avion, la RMSE sur l'altitude à $t = 600$ s est réduite de 50 % à 85 % comparée à la RMSE obtenue avec les paramètres de référence fournis par BADA. L'utilisation de BADA avec les paramètres choisis avec nos méthodes donne de meilleurs résultats que la prédiction avec un modèle reliant directement les variables connues à l'altitude à $t = 600$ s.

Cette thèse s'est concentrée sur l'étude des segments de montée de plus de 10 min car c'était l'horizon de prédiction visé. Ainsi, les modèles sont faits pour prédire des trajectoires de montée provenant de segment de plus de 10 min. Toutefois, suivant le type avion considéré, les segments de moins de 10 min sont majoritaires parmi les segments de montée dans l'espace aérien supérieur. Les paramètres choisis par nos modèles peuvent servir à prédire sur un horizon de moins de 10 min. La prédiction sur des segments plus petits mais d'au moins 5 min est étudiée dans cette thèse. Il ressort de cette étude que la RMSE est bien réduite par nos méthodes sur des petits segments. Cependant, les valeurs de RMSE sont supérieures à celles trouvées en appliquant ces méthodes sur des segments de plus de 10 min. Il est difficile d'expliquer cet écart qui peut avoir plusieurs causes. Il peut provenir du fait que les avions avec un même passé opèrent différemment suivant la durée de montée prévue (*i.e.* $p_{moins}(y|x) \neq p_{plus}(y|x)$). Cet écart peut également provenir du fait que le passé des trajectoires est distribué différemment dans les deux cas (*i.e.* $p_{moins}(x) \neq p_{plus}(x)$).

Les travaux effectués dans le cadre de cette thèse ont fait l'objet de plusieurs publications et communications ([Alligier 11, Alligier 12, Alligier 13b, Alligier 13a, Alligier 14]).

Perspectives

Les méthodes de clustering qui partitionnent un jeu de trajectoires en sous-ensembles de trajectoires similaires n'ont pas été explorées dans cette thèse. Le plus délicat dans ces méthodes est de définir une notion de similarité pertinente pour l'exploitation que l'on veut en faire. Dans ce cadre, une notion de distance entre une commande de poussée et une trajectoire observée ressort de ces travaux. Celle-ci pourrait être exploitée pour obtenir un partitionnement du jeu de trajectoires pour lequel chaque partie contient des trajectoires avec une même commande de poussée. Pour un usage prédictif de cette partition, il suffirait d'estimer un modèle fournissant la commande de poussée à utiliser (*i.e.* la partition à laquelle appartiendrait la trajectoire) en fonction du passé de la trajectoire considérée.

Toutes les trajectoires observées dans l'espace aérien supérieur ne suivent pas un profil de vitesse de type $(cas, Mach)$. Certaines trajectoires ont un profil de vitesse composé d'un segment à $cas1$ constant, puis d'une accélération jusqu'à un profil de type $(cas2, Mach)$. Être capable de prédire ce type de profil pourrait résulter en une amélioration de la prédiction de trajectoire. On pourrait également envisager d'apprendre un profil de vitesse air (TAS), plutôt qu'un profil de vitesse indiquée (cas,Mach).

Les différents cas étudiés confirment l'intérêt des méthodes développées dans cette thèse pour améliorer la prédiction des segments de montée dans l'espace aérien supérieur. Toutefois, tous les cas n'ont pas été étudiés. Il reste à étudier les cas où les 10 derniers points de la trajectoire passée ne correspondent pas à une phase de montée mais à un palier par exemple. Ce cas pourrait être traité en cherchant la phase de montée la plus récente dans le passé et en appliquant nos méthodes sur cette phase. Une fois les paramètres estimés, on pourrait par exemple prendre en compte la consommation de carburant pour recaler la masse estimée quelques minutes avant. Une seconde approche pourrait être d'appliquer les méthodes de cette thèse sur la phase de palier qui précède la montée. Intuitivement, cette seconde approche sera la plus performante pour les valeurs $(cas, Mach)$ et la moins performante pour la masse.

Les méthodes d'amélioration de la prévision de trajectoire introduites dans cette thèse sont développées dans le cadre d'une future application à la détection et résolution de conflits. Certains algorithmes de résolutions de conflits ([Durand 96, Granger 02]) impliquent de tester des trajectoires alternatives i.e. de tester l'impact d'ordres donnés par le contrôleur sur les conflits. Dans ce cadre, suivant le changement de route donné à l'avion, son orientation par rapport au vent sera différente et donc l'effet du gradient de vent sur le taux de montée sera lui aussi différent. Le profil vertical prédit sera différent suivant l'ordre latéral donné. Ce changement de profil vertical suivant l'ordre donné doit être pris en compte dans la prédiction. Ceci est délicat avec un modèle purement statistique tandis qu'un modèle physique le prend naturellement en compte. Plus généralement, il faut savoir, suivant l'ordre donné, quelles informations inférées sur le passé restent pertinentes pour prédire la trajectoire. Ainsi, si l'on donne un ordre intercalant un palier dans une montée, le profil de vitesse et la masse inférés peuvent ne plus être pertinents pour les deux sous-segments de montée résultants. Cette question de l'interaction entre les ordres et les choix sur la mise en œuvre de l'appareil doit être étudiée. Toutefois, on peut supposer qu'utiliser les mêmes paramètres inférés, indépendamment des ordres envisagés, donne une prédiction toujours meilleure que celle obtenue avec les paramètres de référence BADA.

Appendices

Annexe A

Preuve d'existence et de régularité de la fonction $m^*(C_1, \ldots, C_n)$

Dans cette annexe et sous certaines hypothèses, on prouve l'existence de $m^*(C_1, \ldots, C_n)$. On démontre également la régularité de la fonction m^*.

Pour ce faire, on réécrit $\mathcal{E}((C_1, \ldots, C_n), m)$:

$$\begin{aligned}
\mathcal{E}((C_1, \ldots, C_n), m) =& \sum_{i=1}^{n} \left(\frac{P_i(C_i, m + \delta_i)}{m + \delta_{avg}} \right)^2 \\
=& \sum_{i=1}^{n} \left[\frac{\alpha_i(C_i) + (m + \delta_i)\beta_i + (m + \delta_i)^2 \gamma_i}{m + \delta_{avg}} \right]^2 \\
=& \frac{1}{(m + \delta_{avg})^2} \sum_{i=1}^{n} \left[\alpha_i(C_i) + (m + \delta_i)\beta_i + (m + \delta_i)^2 \gamma_i \right]^2 \\
=& \frac{a + (m + \delta_{avg})b + (m + \delta_{avg})^2 c + (m + \delta_{avg})^3 d + (m + \delta_{k,avg})^4 e}{(m + \delta_{avg})^2} \\
=& \frac{a}{(m + \delta_{avg})^2} + \frac{b}{(m + \delta_{avg})} + c + (m + \delta_{avg})d + (m + \delta_{avg})^2 e,
\end{aligned}$$

avec :

$$
\begin{aligned}
Drag_i =& d_{0i} + d_{2i} \times m^2 \\
\alpha_i =& (C_i Thr_i - d_{0i}) tas_i \\
\beta_i =& -Q_i \\
\gamma_i =& -d_{2i} \times tas_i \\
a =& \sum_{i=1}^{n} \left(\alpha_i + \beta_i (\delta_i - \delta_{avg}) + \gamma_i (\delta_i - \delta_{avg})^2\right)^2 \\
b =& 2\sum_{i=1}^{n} \left(\alpha_i + \beta_i (\delta_i - \delta_{avg}) + \gamma_i (\delta_i - \delta_{avg})^2\right) (\beta_i + 2\gamma_i (\delta_i - \delta_{avg})) \\
c =& \sum_{i=1}^{n} (\beta_i + 2\gamma_i (\delta_i - \delta_{avg}))^2 + 2\left(\alpha_i + \beta_i (\delta_i - \delta_{avg}) + \gamma_i (\delta_i - \delta_{avg})^2\right)\gamma_i \\
d =& \sum_{i=1}^{n} (\beta_i + 2\gamma_i (\delta_i - \delta_{avg}))\gamma_i \\
e =& \sum_{i=1}^{n} \gamma_i^2
\end{aligned}
$$

Avec cette réécriture et sans hypothèses supplémentaires, on a : $a \geqslant 0$ et $e \geqslant 0$. Toutefois, les avions ayant une vitesse non nulle par rapport au vent, on a $e > 0$. On suppose $d > 0$, cette hypothèse est validée expérimentalement sur nos données. La valeur de d est indépendante du profil $(C_1, \ldots, C_n)$. À contrario, la valeur de a dépend du profil de $(C_1, \ldots, C_n)$. Il existe d'ailleurs un unique profil tel que $a = 0$. Celui-ci est obtenu en résolvant le système linéaire $(C_i Thr_i - d_{0i}) tas_i + \beta_i (\delta_i - \delta_{avg}) + \gamma_i (\delta_i - \delta_{avg})^2 = 0$. Un tel profil existe toujours et est unique, car la matrice associée à ce système est diagonale avec les termes diagonaux non nuls. Pour la suite, on travaillera sur l'ouvert W que l'on définit comme étant $\mathbb{R}^n$ privé du profil solution du système. Ceci garantit $a > 0$.

A.1 Existence de la fonction $m^*(C_1, \ldots, C_n)$

Pour prouver l'existence de la fonction m^*, on démontre l'existence et l'unicité du minimum global de $\mathcal{E}((C_1, \ldots, C_n), .)$ sur $]-\delta_{avg}; +\infty[$.

A.1.1 Existence d'un minimum global de $\mathcal{E}((C_1, \ldots, C_n), .)$

Pour tout profil dans W :

- $\lim\limits_{m \to -\delta_{avg}^+} \mathcal{E}((C_1, \ldots, C_n), m) = +\infty$, car $a > 0$
- $\lim\limits_{m \to +\infty} \mathcal{E}((C_1, \ldots, C_n), m) = +\infty$, car $e > 0$

Ceci permet de conclure sur l'existence d'un minimum global sur $]-\delta_{avg}; +\infty[$ car $\mathcal{E}((C_1, \ldots, C_n), .)$ est continue.

A.1.2 Unicité du minimum global

Comme $\mathcal{E}((C_1,\dots,C_n),.)$ est dérivable, un minimum global sur $]-\delta_{avg};+\infty[$ ouvert annule $\frac{\partial \mathcal{E}}{\partial m}((C_1,\dots,C_n),.)$. On va montrer qu'il n'y a qu'une seule masse annulant cette fonction.

$$\begin{aligned}
&\frac{\partial \mathcal{E}}{\partial m}((C_1,\dots,C_n),m)=0\\
\Leftrightarrow\ &-2\frac{a}{(m+\delta_{avg})^3}-\frac{b}{(m+\delta_{avg})^2}+d+2e(m+\delta_{avg})=0\\
\Leftrightarrow\ &2e(m+\delta_{avg})^4+d(m+\delta_{avg})^3-b(m+\delta_{avg})-2a=0
\end{aligned}$$

Posons $P(X)=2eX^4+dX^3-bX-2a$, on va maintenant étudier le nombre de racine de P dans $\mathbb{R}^{+*}$.

Une seule racine de P dans $\mathbb{R}^{+*}$

Pour conclure sur le nombre de racine de P dans $\mathbb{R}^{+*}$, on remarque :

1. $P(0)=-2a<0$
2. P' est croissante sur $\mathbb{R}^{+*}$ car
   $\forall x\in\mathbb{R}^{+*},\ P''(x)=\underbrace{24e}_{>0}x^2+\underbrace{6d}_{\geqslant 0}x>0$

Ainsi, on a deux cas assez similaires :
- Si $P'(0)<0$, alors P est décroissante puis croissante mais comme $P(0)<0$ alors P ne coupe l'axe des x qu'une fois
- Si $P'(0)>0$, alors P est croissante et remarquer que $P(0)<0$ permet de conclure

Au final, une seule masse dans $]-\delta_{avg};+\infty[$ peut annuler $\frac{\partial \mathcal{E}}{\partial m}((C_1,\dots,C_n),.)$ ce qui permet de conclure sur l'unicité du minimum global.

A.2 Régularité de la fonction $m^*(C_1,\dots,C_n)$

Pour démontrer la régularité de la fonction m^*, on va appliquer le théorème des fonctions implicites à l'équation $\frac{\partial \mathcal{E}}{\partial m}((C_1,\dots,C_n),.)$ dont le minimum global est l'*unique* solution.

A.2.1 Théorème des fonctions implicites

Notations

Soit $E=\mathbb{R}^p$, $F=\mathbb{R}^q$, Ω ouvert de $E\times F$ et f une application de Ω dans F.
On a donc $f(x,y)=(f_1(x,y),\dots,f_q(x,y))$ avec les f_i définies sur Ω et à valeurs réelles.

Théorème

On suppose que f est de classe C^k et vérifie les hypothèses :

1. $f(a,b) = 0$ avec $a = (a_1, \ldots, a_p)$, $b = (b_1, \ldots, b_q)$ et $(a,b) \in \Omega$
2. $det[\frac{\partial f_i}{\partial y_j}(a,b)]_{1 \leq i,j \leq q} \neq 0$

Alors il existe un ouvert U en a, un ouvert V en b, et une application ϕ de U dans V de classe C^k telle que

1. $U \times V \subset \Omega$
2. $\forall (x,y) \in U \times V, (f(x,y) = 0 \Leftrightarrow y = \phi(x))$

A.2.2 Application du théorème

En utilisant les notations introduites dans A.2.1, on va appliquer le théorème des fonctions implicites avec :

- $f = \frac{\partial \mathcal{E}}{\partial m}$
- $p = n$ et $q = 1$
- $E = \mathbb{R}^n$ et $F = \mathbb{R}$
- $\Omega = W \times]-\delta_{avg}; +\infty[$ ouvert
- $a = (C_1, \ldots, C_n)$ et $b = m^* (C_1, \ldots, C_n)$

La fonction f est de classe C^∞

Pour prouver que f est de classe C^∞ on va prouver que toutes ses dérivées partielles d'ordres quelconques sont définies et continues sur Ω.

On remarque d'abord que f est une fonction rationnelle et peut s'écrire $\frac{u}{v}$. Sur tout Ω, v ne s'annule pas. Ainsi les dérivées partielles existent et sont de la forme $\frac{u'v - uv'}{v^2}$. Les dérivées partielles obtenues sont donc des fonctions rationnelles. Elles sont continues sur leur ensemble de définition c'est-à-dire Ω. On peut appliquer de nouveau ce principe à chacune des dérivées partielles d'ordres supérieures.

On a donc prouvé que toutes les dérivées partielles d'ordres quelconques sont définies et continues sur Ω ce qui permet de conclure que f est de classe C^∞.

Si $f((C_1, \ldots, C_n), m) = 0$ **alors** $\frac{\partial f}{\partial m}((C_1, \ldots, C_n), m) \neq 0$

Pour tout $\theta \in \mathbb{R}^n$ et tout $m \in \mathbb{R}^{+*}$, on a :

$$\frac{\partial f}{\partial m}(\theta, m) = 6\frac{a}{m^4} + 2\frac{b}{m^3} + 2e = \frac{2em^4 + 2bm + 6a}{m^4}$$

Si l'on suppose que m soit solution de $\frac{\partial f}{\partial m}(\theta, m) = 0$, on a $P(m) = 2em^4 + dm^3 - bm - 2a = 0$ et donc $bm = 2em^4 + dm^3 - 2a$

Ainsi, en remplaçant Bm dans la dérivée seconde, on a :

$$\frac{\partial f}{\partial m}(\theta, m) = \frac{2em^4 + 2(2em^4 + dm^3 - 2a) + 6a}{m^4} = \frac{6em^4 + 2dm^3 + 2a}{m^4} > 0$$

Conclusion

En appliquant le théorème des fonctions implicites, on a ϕ de classe C^k, U et V tels que :

$$\forall \left((C_1, \ldots, C_n), m \right) \in U \times V, f\left((C_1, \ldots, C_n), m \right) = 0 \Leftrightarrow m = \phi\left(C_1, \ldots, C_n \right) \qquad \text{(A.1)}$$

Dans la section A.1.2, on a montré que pour tout $(C_1, \ldots, C_n)$ dans W, $m^*\left(C_1, \ldots, C_n \right)$ est l'*unique* valeur annulant $f((C_1, \ldots, C_n), .)$ ce qui permet de déduire que $m^* = \phi$ sur U.

On a ainsi démontré que pour toute valeur de $a = (C_1, \ldots, C_n)$ il existe un voisinage U de a et une fonction ϕ de classe C^k tels que $m^* = \phi$ sur ce voisinage. Comme ceci s'applique pour tout a dans W, on prouve ainsi que m^* de classe C^k sur tout W. Ceci étant vrai pour k quelconque, m^* est de classe C^∞.

Annexe B

Ajustement du profil (CAS,$Mach$)

Lors de la phase de montée, selon le manuel d'utilisateur BADA ([Nuic 11]), l'avion suit un profil de vitesse paramétré par une valeur de CAS *cas* et une valeur de Mach M. L'avion vole à une CAS égale à *cas* en dessous de l'altitude de conjonction $H_{p,trans}$ et à Mach égal à M au dessus. La définition de l'altitude de conjonction est cohérente avec la définition de ce profil de vitesse : c'est l'altitude à laquelle la valeur de CAS *cas* et de Mach M font référence à la même TAS. Cette altitude pression dépend uniquement du couple (*cas*, M) comme on peut le voir dans la formule (B.1).

$$H_{p,trans} = \frac{1000}{0.30488 * 6.5} \left[T_0 \left(1 - \theta_{trans}\right)\right] \tag{B.1a}$$

$$\text{avec } \theta_{trans} = \left(\delta_{trans}\right)^{\frac{-\beta_T R}{g_0}} \tag{B.1b}$$

$$\text{et } \delta_{trans} = \frac{\left[1 + \frac{\kappa-1}{2}\left(\frac{cas}{a_0}\right)^2\right]^{\frac{\kappa}{\kappa-1}} - 1}{\left[1 + \frac{\kappa-1}{2}M^2\right]^{\frac{\kappa}{\kappa-1}} - 1} \tag{B.1c}$$

$$\text{et } T_0\text{, la température ISA au niveau de la mer} \tag{B.1d}$$

$$\text{et } a_0\text{, la célérité du son dans l'air au niveau de la mer} \tag{B.1e}$$

$$\text{et } \beta_T\text{, le gradient de température} \tag{B.1f}$$

$$\text{et } R\text{, la constante universelle des gaz parfaits} \tag{B.1g}$$

$$\text{et } g_0\text{, l'accélération de la pesanteur à la surface de la Terre} \tag{B.1h}$$

$$\text{et } \kappa\text{, l'indice adiabatique de l'air} \tag{B.1i}$$

En lien avec la définition de l'altitude pression, sachant la pression p_{trans} de l'altitude de conjonction, les deux formules (B.2) relient la CAS et le Mach.

$$CAS_{trans}(p_{trans}, M) = \left[\frac{2}{\mu} RT_0 \left\{ \left(1 + \frac{p_{trans}}{p_0} \left[\left(1 + \frac{\mu}{2}\kappa M^2 \right)^{\frac{1}{\mu}} - 1 \right] \right)^{\mu} - 1 \right\} \right]^{\frac{1}{2}} \tag{B.2a}$$

$$\text{avec } \mu = \frac{\kappa - 1}{\kappa} \tag{B.2b}$$

$$\text{et } p_0, \text{ la pression ISA au niveau de la mer} \tag{B.2c}$$

Lorsque (cas,M) est connue, à une altitude pression et une température donnée, la TAS peut être calculée par les formules (B.3) ci-dessous.

$$TAS(cas, M, p, T) = \begin{cases} f(cas, p, T) & \text{si } H_p(p) \leqslant H_{p,trans}(cas, M) \\ f(cas, p_{trans}, T) & \text{sinon} \end{cases} \tag{B.3a}$$

$$\text{avec } f(cas, p, T) = \left[\frac{2}{\mu} RT \left\{ \left(1 + \frac{p_0}{p} \left[\left(1 + \frac{\mu}{2}\frac{\rho_0}{p_0} cas^2 \right)^{\frac{1}{\mu}} - 1 \right] \right)^{\mu} - 1 \right\} \right]^{\frac{1}{2}} \tag{B.3b}$$

Le choix du paramétrage de ce profil de vitesse est fait par le pilote. Il peut être le résultat de l'optimisation du coût de la mise en œuvre de l'appareil. Classiquement, ce coût est modélisé comme la somme d'un coût fixe, d'un coût proportionnel au temps et d'un coût proportionnel à la quantité de carburant consommé.

Le profil de vitesse est une donnée très importante pour prédire l'altitude future d'un avion. En effet, la puissance est partagée entre la variation d'énergie cinétique et la variation d'énergie potentielle. Une fois que l'on soustrait la variation d'énergie cinétique nécessaire pour suivre le profil de vitesse, la puissance restante est utilisée pour augmenter l'énergie potentielle et donc l'altitude.

B.1 Le problème d'optimisation associé

Au regard des profils de vitesse de notre jeu de trajectoires réelles, certains avions ne suivent visiblement pas un profil de vitesse (CAS,Mach). Dans cette section, on se propose d'extraire un couple (cas, M) qui pourrait correspondre à la trajectoire observée constituée de n points. On pourra par la suite distinguer les avions qui ont suivi un profil de vitesse (CAS,Mach) des autres en regardant l'écart entre le profil ajusté (cas, M) et celui observé.

Cet écart peut être quantifié par la somme des différences au carré. On va ajuster notre profil de vitesse pour minimiser Φ définie par l'équation (B.4).

$$\Phi(cas, M) = \sum_{i=1}^{n} \left(TAS(cas, M, p_i, T_i) - TAS_i \right)^2 \tag{B.4a}$$

$$= \sum_{i \in [\![1;n]\!] / H_{p_i} \leqslant H_{p,trans}(cas,M)} \left(f(cas, p_i, T_i) - TAS_i \right)^2 \tag{B.4b}$$

$$+ \sum_{i \in [\![1;n]\!] / H_{p,trans}(cas,M) < H_{p_i}} \left(f(cas, p_{trans}, T_i) - TAS_i \right)^2 \tag{B.4c}$$

La fonction à minimiser Φ est continue. Du fait de la définition (B.3) de la fonction TAS, Φ n'est pas différentiable aux points (cas, M) vérifiant $H_p(p_i) = H_{p,trans}(cas, M)$. Φ peut s'exprimer comme somme de fonctions $\psi_{(p,T,TAS)}(cas) = (f(cas, p, T) - TAS)^2$.

Comme Φ n'est pas différentiable sur tout le domaine, sans redécoupage préalable, il faut donc utiliser une méthode n'utilisant pas la différentielle. Parmi les méthodes d'optimisation globale que l'on pourrait utiliser, on peut citer les méta-heuristiques telles que le recuit simulé ([Kirkpatrick 83]), les algorithmes génétiques ([Goldberg 89]) ou l'évolution différentielle ([Storn 97]). Ces méthodes ont l'avantage de pouvoir converger vers le minimum global. Toutefois le critère d'arrêt de ces algorithmes est bien souvent fixé de manière arbitraire. Il est difficile de quantifier à quel point on est proche d'un minimum global.

On se propose de découper le domaine en sous-domaines. La résolution sur chacun de ces sous-domaines est bien plus aisée. En effet, cette résolution se réduit à trouver $x \in I$ annulant une fonction g, avec I un intervalle réel. En supposant g monotone sur I, la recherche dichotomique est une méthode simple et efficace qui permet d'encadrer la solution x^*. C'est une méthode qui maintient un encadrement de x^* qui converge en $O(2^n)$, avec une seule évaluation de g à chaque itération. C'est donc une méthode rapide qui permet de fixer un critère d'arrêt portant sur l'encadrement de la solution. Dans ce cadre, on va en particulier s'intéresser aux fonctions $\psi_{(p,T,TAS)}$.

B.1.1 Preuve de convexité de $\psi_{(p,T,TAS)}$ sur I_{cas}

Pour démontrer la convexité de $\psi_{(p,T,TAS)}$, on va utiliser l'arithmétique des intervalles introduite dans [Moore 66] par Moore pour prouver que ${\psi''}_{(p,T,TAS)}$ est positif. Dans l'arithmétique par intervalle, on associe à chaque opérateur portant sur les réels un opérateur équivalent portant sur des intervalles. Un opérateur $\tilde{op}$ associé à l'opérateur op sur $\mathbb{R}^n$ satisfait la propriété d'inclusion suivante : pour tout produit d'intervalles $\prod_{i=1}^{n}[a_i; b_i]$, $\forall x \in \prod_{i=1}^{n}[a_i; b_i],\ op(x) \in \tilde{op}(\prod_{i=1}^{n}[a_i; b_i])$. L'ensemble retourné par $\tilde{op}$ est lui même un intervalle, ce qui va permettre de pouvoir composer entre eux les opérateurs de l'arithmétique par intervalle. Ainsi, en décomposant une fonction quelconque en composition d'opérateurs ayant un équivalent dans l'arithmétique par intervalle, l'arithmétique par intervalle permet de borner la fonction considérée.

En utilisant l'arithmétique par intervalle, on montre que la dérivé seconde de $\psi_{(p,T,TAS)}$ est positive sur le domaine défini dans l'équation (B.5).

$$\begin{gathered} I_p = [5474.9; 110000], \text{ en Pa} \\ I_T = [150; 350], \text{ en K} \\ I_{TAS} = [25; 500], \text{ en m/s} \\ I_{cas} = [5; 750], \text{ en m/s} \\ {\psi''}_{(I_p,I_T,I_{TAS})}(I_{cas}) \subset [2.804021e-07; 7.711511e+01] \end{gathered} \tag{B.5}$$

Ce domaine contient tout les triplets (p, T, TAS) observés dans notre jeu de trajectoires. Ainsi, $\forall (p, T, TAS) \in I_p \times I_T \times I_{TAS}$, $\psi_{(p,T,TAS)}$ convexe sur I_{cas}.

On va maintenant montrer que le minimum sur $\mathbb{R}^{+*}$ d'une fonction constituée d'une somme de fonctions de type $\psi_{(p,T,TAS)}$ se situe sur I_{cas}.

$$\psi'_{(p,T,TAS)}(cas) = 2\left(f(cas,p,T) - TAS\right)\underbrace{\frac{\partial f}{\partial cas}(cas,p,T)}_{>0} \tag{B.6}$$

Avec l'équation ci-dessus, on a $\psi'_{(p,T,TAS)}$ du même signe que $(f(cas,p,T) - TAS)$. On va donc s'intéresser aux valeurs prises par $f(cas,p,T)$.

$$f(\{5\}, I_p, I_T) \subset [3.462336, 23.695177] \tag{B.7a}$$
$$f(\{750\}, I_p, I_T) \subset [529.478768; 1550.998766] \tag{B.7b}$$

Avec les équations ci-dessus, et en se rappelant que $\forall (p,T) \in I_p \times I_T,\ f(.,p,T)$ est croissante, on peut conclure sur le signe de la dérivée des fonctions $\psi_{(p,T,TAS)}$.

$$\forall (p,T,TAS) \in I_p \times I_T \times I_{TAS},\ \forall x \in]0;5[,\ \psi'_{(p,T,TAS)}(x) < 0 \tag{B.8a}$$
$$\forall (p,T,TAS) \in I_p \times I_T \times I_{TAS},\ \forall x \in]750;+\infty[,\ \psi'_{(p,T,TAS)}(x) > 0 \tag{B.8b}$$

Ainsi, $\forall (p,T,TAS) \in I_p \times I_T \times I_{TAS}$, $\psi_{(p,T,TAS)}$ est décroissante sur $]0;5[$ et croissante sur $]750;+\infty[$. Par conséquent, la dérivée d'une fonction constitué d'une somme de fonctions de type $\psi_{(p,T,TAS)}$ ne peut s'annuler que sur I_{cas}. Ainsi lorsque l'on va chercher le point d'annulation de la dérivée de la fonction, on va pouvoir restreindre cette recherche sur I_{cas}.

B.2 Un découpage en sous-domaines

Cette non différentiabilité de Φ provient du fait que la fonction TAS a une coupure en $H_p(p) = H_{p,trans}(cas, M)$. On va donc découper le domaine $\Omega = \mathbb{R}^{+*} \times \mathbb{R}^{+*}$ en sous-domaines sur lesquels Φ est différentiable.

$$\Omega = O_M \cup O_{cas} \cup \bigcup_{k=1}^{n-1} O_k \cup \bigcup_{k=1}^{n} F_k \tag{B.9a}$$
$$\text{avec } O_M = {H_{p,trans}}^{-1}\left(]0; H_p(p_1)[\right) \tag{B.9b}$$
$$O_{cas} = {H_{p,trans}}^{-1}\left(]H_p(p_n); +\infty[\right) \tag{B.9c}$$
$$O_k = {H_{p,trans}}^{-1}\left(]H_p(p_k); H_p(p_{k+1})[\right) \tag{B.9d}$$
$$F_k = {H_{p,trans}}^{-1}\left(H_p(p_k)\right) \tag{B.9e}$$

B.2.1 Résolution sur O_M

Sur O_M, le profil de vitesse est indépendant de cas. On cherche à trouver M minimisant Φ_{O_M} tel que défini par l'équation (B.10).

$$\Phi_{O_M}(M) = \sum_{i=1}^{n} \left(M\sqrt{\kappa R T_i} - TAS_i\right)^2 \tag{B.10}$$

La fonction Φ_{O_M} est évidemment convexe et son minimum global sur O_M ouvert, si il existe, annule donc la dérivée.

$$\frac{\partial \Phi_{O_M}}{\partial M}(M) = 0 \tag{B.11a}$$

$$\Leftrightarrow \sum_{i=1}^{n} \left(M\sqrt{\kappa RT_i} - TAS_i \right) \sqrt{\kappa RT_i} = 0 \tag{B.11b}$$

$$\Leftrightarrow M = \frac{\sum_{i=1}^{n} TAS_i \sqrt{\kappa RT_i}}{\sum_{i=1}^{n} \kappa RT_i} \tag{B.11c}$$

On connait donc le minimum de Φ_{O_M} sur $\mathbb{R}^{+*}$.

B.2.2 Résolution sur O_{cas}

Sur O_{cas}, le profil de vitesse est indépendant de M. On cherche à trouver cas minimisant $\Phi_{O_{cas}}$ tel que défini par l'équation (B.12).

$$\Phi_{O_{cas}}(cas) = \sum_{i=1}^{n} \psi_{(p_i, T_i, TAS_i)}(cas) \tag{B.12}$$

Comme l'on recherche le minimum global sur $\mathbb{R}^{+*}$ ouvert ; si celui-ci existe, il annule la dérivée. Or, elle ne peut s'annuler que sur I_{cas}. On va donc chercher l'annulation dans I_{cas}. De plus $\Phi_{O_{cas}}$ est convexe sur I_{cas}, on va donc pouvoir utiliser une recherche dichotomique sur l'intervalle I_{cas}.

B.2.3 Résolution sur O_k

Comme $H_{p,trans}$ est continue, on a O_k ouvert. Ainsi, si (cas, M) est minimum sur O_k alors $\nabla \Phi(cas, M) = 0$.

En particulier, on a :

$$\frac{\partial \Phi}{\partial M}(cas, M) = 2 \sum_{i=k+1}^{n} \left(M\sqrt{\kappa RT_i} - TAS_i \right) \sqrt{\kappa RT_i} = 0 \tag{B.13a}$$

$$\Leftrightarrow M = \frac{\sum_{i=k+1}^{n} TAS_i \sqrt{\kappa RT_i}}{\sum_{i=k+1}^{n} \kappa RT_i} \tag{B.13b}$$

On sait donc que nécessairement, $M = \frac{\sum_{i=k+1}^{n} TAS_i\sqrt{\kappa RT_i}}{\sum_{i=k+1}^{n} \kappa RT_i}$. Il nous reste à trouver cas dans O_k qui annule $\frac{\partial\Phi}{\partial cas}(.,M)$. Or, on peut caractériser l'appartenance de (cas, M) à O_k.

$$\begin{array}{cc} & (cas, M) \in O_k \\ \Leftrightarrow & H_p(p_k) < H_{p,trans}(cas, M) < H_p(p_{k+1}) \\ \Leftrightarrow & p_{k+1} < p_{trans} < p_k \\ \underset{CAS_{trans}(.,M) \text{ croissante}}{\Leftrightarrow} & CAS_{trans}(p_{k+1}, M) < cas < CAS_{trans}(p_k, M) \end{array} \tag{B.14}$$

La seconde dérivée à annuler est donnée par l'équation (B.15) ci-dessous.

$$\frac{\partial\Phi}{\partial cas}(cas, M^*) = \sum_{i=1}^{k} \psi'_{(p_i,T_i,TAS_i)}(cas) \tag{B.15}$$

On cherche où s'annule la dérivée de $cas \longmapsto \Phi(cas, M^*)$ dans $]CAS_{trans}(p_{k+1}, M^*); CAS_{trans}(p_k, M^*)$ Or cette fonction ne peut s'annuler que sur I_{cas}. On va donc chercher l'annulation dans $I_{cas} \cap]CAS_{trans}(p_{k+1}, M^*); CAS_{trans}(p_k, M^*)[$. De plus $cas \longmapsto \Phi(cas, M^*)$ est convexe sur I_{cas}, on va donc pouvoir utiliser une recherche dichotomique sur l'intervalle $I_{cas} \cap]CAS_{trans}(p_{k+1}, M^*); CA$

B.2.4 Résolution sur F_k

L'ensemble F_k est caractérisé par le fait que $p_{trans} = p_k$. Ainsi avec les formules (B.2), on a $M = M_{trans}(p_k, cas)$.

$$\Phi_{F_k}(cas) = \Phi(cas, M_{trans}(p_k, cas)) \tag{B.16a}$$

$$= \sum_{i=1}^{k} \psi_{(p_i,T_i,TAS_i)}(cas) + \sum_{i=k+1}^{n} \psi_{(p_k,T_i,TAS_i)}(cas) \tag{B.16b}$$

Comme l'on recherche le minimum global de Φ_{F_k} sur $\mathbb{R}^{+*}$ ouvert ; si celui-ci existe, il annule la dérivée de Φ_{F_k}. Or, elle ne peut s'annuler que sur I_{cas}. On va donc chercher l'annulation dans I_{cas}. De plus Φ_{F_k} est convexe sur I_{cas}, on va donc pouvoir utiliser une recherche dichotomique sur l'intervalle I_{cas}.

B.2.5 Conclusion sur le domaine Ω

Le minimum de Φ sur Ω appartient à un des sous-domaines O_M, O_{cas},F_k ou O_k et est aussi minimum de ce sous-domaine. Ainsi, si l'on isole les minimums possibles de chacun de ces sous-domaines, le minimum sur Ω sera parmi ces minimum possibles. Ainsi, pour conclure sur le minimum sur Ω, il suffit de sélectionner celui minimisant Φ parmi les minimum des sous-domaines.

Glossaire

AMAN Arrival MANager.
ATM Air Traffic Management.

BADA Base of Aircraft DAta.

CAS Calibrated AirSpeed.
CTR Control Traffic Region.

ÉNAC École Nationale de l'Aviation Civile.
ESF Energy Share Factor.

FABEC Functional Airspace Block Europe Central.
FL Flight Level.
FMS Flight Management System.

IAS Indicated AirSpeed.
IFR Instrument Flight Rules.
ISA International Standard Atmosphere.

NM Nautical Mile.

RMSE Root Mean Square Error.

SEP Specific Excess Power.
SER Specific Energy Rate.
SESAR Single European Sky ATM Research.

TAS True AirSpeed.
TMA Terminal Manœuvring Area.

VFR Visual Flight Rules.

Bibliographie

[Akaike 74] H. Akaike. *A new look at the statistical model identification.* Automatic Control, IEEE Transactions on, vol. 19, no. 6, pages 716–723, 1974. *Cité p.* 42

[Alligier 11] R. Alligier, M. Ghasemi Hamed, D. Gianazza et M. Serrurier. *Ground-based prediction of aircraft climb : point-mass model vs regression methods.* Complex World 2011, 1st Annual Complex World Conference, Seville, Spain, juillet 2011. *Cité p.* 3, 186

[Alligier 12] R. Alligier, D. Gianazza et N. Durand. *Energy Rate Prediction Using an Equivalent Thrust Setting Profile (regular paper).* International Conference on Research in Air Transportation (ICRAT), Berkeley, California, 22/05/12-25/05/12, page (on line), http ://www.icrat.org, may 2012. ICRAT. *Cité p.* 3, 97, 186

[Alligier 13a] R. Alligier, D. Gianazza et N. Durand. *Ground-based Estimation of Aircraft Mass, Adaptive vs. Least Squares Method.* 10th USA/Europe Air Traffic Management Research and Developpment Seminar, 2013. *Cité p.* 3, 100, 186

[Alligier 13b] R. Alligier, D. Gianazza et N. Durand. *Learning the aircraft mass and thrust to improve the ground-based trajectory prediction of climbing flights.* Transportation Research Part C : Emerging Technologies, vol. 36, no. 0, pages 45 – 60, 2013. *Cité p.* 3, 138, 186

[Alligier 14] R. Alligier, M. Ghasemi Hamed, D. Gianazza et N. Durand. *Comparison of Two Ground-based Mass Estimation Methods on Real Data (regular paper).* International Conference on Research in Air Transportation (ICRAT), Istanbul, Turkey, 26/05/14-31/05/14, page (on line), http ://www.icrat.org, may 2014. ICRAT. *Cité p.* 3, 186

[Amaldi 98] E. Amaldi et V. Kann. *On the approximability of minimizing nonzero variables or unsatisfied relations in linear*

systems. Theoretical Computer Science, vol. 209, no. 1, pages 237–260, 1998. *Cité p.* 44

[Bishop 95a] C. M. Bishop. Neural networks for pattern recognition. Oxford University Press, Inc., New York, NY, USA, 1995. *Cité p.* 52, 54

[Bishop 95b] C. M. Bishop. *Training with noise is equivalent to Tikhonov regularization*. Neural Comput., vol. 7, no. 1, pages 108–116, janvier 1995. *Cité p.* 53

[Bishop 06] C. M Bishop. Pattern recognition and machine learning, volume 1. springer New York, 2006. *Cité p.* 36

[Breiman 84] L. Breiman, J. H. Friedman, R. A. Olshen et C. J. Stone. Classification and regression trees. Statistics/Probability Series. Wadsworth Publishing Company, Belmont, California, U.S.A., 1984. *Cité p.* 55, 56, 57

[Broyden 70] C. G. Broyden. *The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations*. IMA Journal of Applied Mathematics, vol. 6, no. 1, pages 76–90, 1970. *Cité p.* 52, 142

[Bühlmann 07] P. Bühlmann et T. Hothorn. *Boosting algorithms : Regularization, prediction and model fitting*. Statistical Science, pages 477–505, 2007. *Cité p.* 57

[Buja 01] A. Buja et Y.-S. Lee. *Data mining criteria for tree-based regression and classification*. Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '01, pages 27–36, New York, NY, USA, 2001. ACM. *Cité p.* 56

[Calders 02] P. Calders. *G.A.M.E. Aircraft Performance Model Description*. Rapport technique 1.0, Eurocontrol, septembre 2002. *Cité p.* 16

[Cornuéjols 10] A. Cornuéjols et L. Miclet. Apprentissage artificiel : Concepts et algorithmes. Eyrolles, juin 2010. *Cité p.* 36

[dos Santos 02] J. M. dos Santos. *Eight-States Free Route Airspace Project : Free Route Airspace Concept*. Rapport technique 2.0, Eurocontrol, Bruxelles, février 2002. *Cité p.* 12

[Duarte Silva 01] A. P. Duarte Silva. *Efficient Variable Screening for Multivariate Analysis*. Journal of Multivariate Analysis, vol. 76, no. 1, pages 35 – 62, 2001. *Cité p.* 44

[Duong 01] V. Duong, G. Gawinowski, J.-P. Nicolaon et D. Smith. *Sector-Less Air Traffic Management*. International Air

Traffic Management R&D Seminar ATM-2001, Sante-Fe, décembre 2001. *Cité p.* 12

[Durand 96] N. Durand. *Optimisation de trajectoires pour la résolution de conflits*. PhD thesis, Thèse doctorat informatique INPT, 1996. *Cité p.* 1, 16, 17, 187

[Edelman 95] A. Edelman et H. Murakami. *Polynomial roots from companion matrix eigenvalues*. Mathematics of Computation, vol. 64, no. 210, pages 763–776, 1995. *Cité p.* 100

[Eurocontrol 97] Eurocontrol. *Eurocontrol Standard Document For Radar Surveillance in En-Route Airspace and Major Terminal Areas*. Rapport technique, Eurocontrol, March 1997. *Cité p.* 104, 109

[Eurocontrol 10] Eurocontrol. *Aircraft Systems and 4D Trajectory Management*. Rapport technique 1.0, Eurocontrol, Bruxelles, septembre 2010. *Cité p.* 13, 15

[Eurocontrol 11] Eurocontrol, Brussels. *As the Crow Flies – Free Route Airspace Maastricht*, mars 2011. *Cité p.* 12

[Eurocontrol 12a] Eurocontrol. *Concept Document for the Base of Aircraft Data (BADA) Family 4*. Rapport technique 1.0, Eurocontrol, novembre 2012. *Cité p.* 14

[Eurocontrol 12b] Eurocontrol. *European Free Route Airspace Developments*. Rapport technique 1.0, Eurocontrol, Bruxelles, février 2012. *Cité p.* 12

[FAA / EUROCONTROL 04] FAA / EUROCONTROL. *Common Trajectory Prediction Capability*. Rapport technique 1.0, FAA / EUROCONTROL COOPERATIVE R & D, avril 2004. *Cité p.* 14

[Fletcher 70] R. Fletcher. *A new approach to variable metric algorithms*. The Computer Journal, vol. 13, no. 3, pages 317–322, 1970. *Cité p.* 52, 142

[Fox 97] J. Fox. Applied regression analysis, linear models, and related methods. Sage Publications, Inc, 1997. *Cité p.* 46

[Freund 97] Y. Freund et R. E. Schapire. *A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting*, 1997. *Cité p.* 55

[Friedman 00] J. H. Friedman. *Greedy Function Approximation : A Gradient Boosting Machine*. Annals of Statistics, vol. 29, pages 1189–1232, 2000. *Cité p.* 45, 55, 57, 58

[Friedman 02] J. H. Friedman. *Stochastic gradient boosting*. Computational Statistics Data Analysis, vol. 38, no. 4, pages 367 – 378, 2002. *Cité p.* 58

[Gallo 07] E. Gallo, J. Lopez-Leones, M. A. Vilaplana, F. A. Navarro et A. Nuic. *Trajectory Computation Infrastructure Based on BADA Aircraft Performance Model*. Digital Avionics Systems Conference, 2007. DASC'07. IEEE/AIAA 26th, pages 1–C. IEEE, 2007. *Cité p.* 33

[Ghasemi Hamed 14] M. Ghasemi Hamed. *Méthodes non-paramétriques pour la prévision d'intervalles avec haut niveau de confiance : application à la prévision de trajectoires d'avions*. PhD thesis, Thèse doctorat informatique de l'INPT, 2014. *Cité p.* 2, 3, 17

[Goldberg 89] D. E. Goldberg. Genetic algorithms in search, optimization, and machine learning. Addison-Wesley Professional, 1 edition, January 1989. *Cité p.* 44, 52, 199

[Goldfarb 70] D. Goldfarb. *A Family of Variable-Metric Methods Derived by Variational Means*. Mathematics of Computation, vol. 24, no. 109, pages pp. 23–26, 1970. *Cité p.* 52, 142

[Granger 02] G. Granger. *Détection et résolution de conflits aériens : modélisations et analyse*. PhD thesis, Thèse doctorat informatique de l'Ecole Polytechnique, 2002. *Cité p.* 1, 16, 17, 18, 187

[Guyon 03] I. Guyon et A. Elisseeff. *An introduction to variable and feature selection*. J. Mach. Learn. Res., vol. 3, pages 1157–1182, mars 2003. *Cité p.* 44, 132

[Hadjaz 12] A. Hadjaz, G. Marceau, P. Savéant et M. Schoenauer. *Online Learning for Ground Trajectory Prediction*. SESAR 2nd Innovation Days, 2012. *Cité p.* 34

[Hairer 93] E. Hairer, S. P. Nørsett et G. Wanner. *Solving ordinary differential equations, I : Nonstiff problems, Springer Verlag*. Berlin, Germany, 1993. *Cité p.* 33

[Hancock 96] T. Hancock, T. Jiang, M. Li et J. Tromp. *Lower Bounds on Learning Decision Lists and Trees*. Information and Computation, vol. 126, no. 2, pages 114 – 122, 1996. *Cité p.* 56

[Hansen 83] E. R. Hansen et R. I. Greenberg. *An interval Newton method*. Applied Mathematics and Computation, vol. 12, no. 2, pages 89–98, 1983. *Cité p.* 101

[Hartley 03] R. Hartley et A. Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003. *Cité p.* 101

[Hastie 01] T. Hastie, R. Tibshirani et J. H. Friedman. The elements of statistical learning. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001. *Cité p.* 36, 41, 43, 57

[Heath 93] D. Heath, S. Kasif et S. Salzberg. *Induction of Oblique Decision Trees*. Journal of Artificial Intelligence Research, vol. 2, no. 2, pages 1–32, 1993. *Cité p.* 56

[Higham 02] N. J. Higham. Accuracy and stability of numerical algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, second edition, 2002. *Cité p.* 100

[Hoerl 70] A. E. Hoerl et R. W. Kennard. *Ridge Regression : Biased Estimation for Nonorthogonal Problems*. Technometrics, vol. 12, no. 1, pages pp. 55–67, 1970. *Cité p.* 40, 42, 46

[Huchet 06] F. Huchet. *Introduction à la prévision de trajectoire : analyse de cas sur 20 vols Paris/Zurich*. Rapport technique, DSNA/DTI/SDER/NT06-832, 2006. *Cité p.* 33

[Jolliffe 82] I. T. Jolliffe. *A Note on the Use of Principal Components in Regression*. Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 31, no. 3, pages pp. 300–303, 1982. *Cité p.* 51

[Kaiser 11] M. Kaiser, M. Schultz et H. Fricke. *Enhanced jet performance model for high precision 4D flight path prediction*. Proceedings of the 1st International Conference on Application and Theory of Automation in Command and Control Systems, ATACCS '11, pages 33–40, Toulouse, France, France, 2011. IRIT Press. *Cité p.* 109

[Kirkpatrick 83] S. Kirkpatrick, C. D. Gelatt et M. P. Vecchi. *Optimization by simulated annealing*. Science, vol. 220, pages 671–680, 1983. *Cité p.* 52, 199

[Kohavi 97] R. Kohavi et G. H. John. *Wrappers for Feature Subset Selection*. ARTIFICIAL INTELLIGENCE, vol. 97, no. 1, pages 273–324, 1997. *Cité p.* 44

[Konyak 08] M. A. Konyak, D. Warburton, J. Lopez-Leones et P. C. Parks. *A Demonstration of an Aircraft Intent Interchange Specification for Facilitating Trajectory-Based Operations in the National Airspace System*. AIAA Guidance, Navigation and Control Conference and Exhibit, Honolulu, Hawaii, 2008. *Cité p.* 17

[Le Fablec 99] Y. Le Fablec. *Prévision de trajectoires d'avions par réseaux de neurones*. PhD thesis, Thèse doctorat informatique de l'INPT, 1999. *Cité p.* 2, 3, 16, 17

[Le Merrer 12] M. Le Merrer. *Optimisation de trajectoire d'avion pour la prise en compte du bruit dans la gestion du vol*. PhD thesis, 2012. *Cité p.* 34

[López-Leonés 07] J. López-Leonés, M. A. Vilaplana, E. Gallo, F. A. Navarro et C. Querejeta. *The aircraft intent description language : a key enabler for air-ground synchronization in trajectory-based operations.* Digital Avionics Systems Conference, 2007. DASC'07. IEEE/AIAA 26th, pages 1–D. IEEE, 2007. *Cité p.* 17, 28

[Lymperopoulos 06] I. Lymperopoulos, J. Lygeros et A. Lecchini Visintini. *Model Based Aircraft Trajectory Prediction during Take-off.* AIAA Guidance, Navigation and Control Conference and Exhibit, Keystone, Colorado, aug 2006. *Cité p.* 2, 92

[MacQueen 67] J. MacQueen. *Some methods for classification and analysis of multivariate observations.* Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1, page 14. California, USA, 1967. *Cité p.* 149

[Martin 71] R. S. Martin, G. Peters et J. H. Wilkinson. *The QR algorithm for real Hessenberg matrices.* Handbook for Automatic Computation, pages 359–371. Springer, 1971. *Cité p.* 100

[Massy 65] W. F. Massy. *Principal Components Regression in Exploratory Statistical Research.* Journal of the American Statistical Association, vol. 60, no. 309, pages pp. 234–256, 1965. *Cité p.* 51

[Mehadhebi 99a] K. Mehadhebi. *Lissage en vitesse.* Rapport technique, CENA, 1999. *Cité p.* 66

[Mehadhebi 99b] K. Mehadhebi. *Lissage horizontal de trajectoires radar par des splines cubiques.* Rapport technique, CENA, 1999. *Cité p.* 66

[Mehadhebi 02] K. Mehadhebi. *Algorithme de spline à rigidité variable, application au lissage de trajectroires radar.* Rapport technique, CENA, 2002. *Cité p.* 66

[Mondoloni 06] S. Mondoloni. *Aircraft Trajectory Prediction Errors : Including a Summary of Error Sources and Data (Version 0.2).* Rapport technique, FAA/Eurocontrol, 2006. *Cité p.* 33

[Moore 66] R. E. Moore. Interval analysis. Prentice-Hall, 1966. *Cité p.* 101, 199

[Mutuel 13] L. H. Mutuel, P. Neri et E. Paricaud. *Initial 4D Trajectory Management Concept Evaluation.* 10th USA/Europe Air Traffic Management Research and Developpment Seminar, 2013. *Cité p.* 13

[Nocedal 06] J. Nocedal et S. J. Wright. Numerical optimization. Springer series in operations research and financial engineering. Springer, New York, NY, 2. ed. edition, 2006. *Cité p.* 52

[Nuic 11] A. Nuic. *User Manual for Base of Aircarft DATA (BADA) Rev.3.9.* Rapport technique, EUROCONTROL, avril 2011. *Cité p.* 14, 28, 31, 84, 129, 197

[Pearson 01] K. Pearson. *On lines and planes of closest fit to systems of points in space.* Philosophical Magazine, vol. 2, no. 6, pages 559–572, 1901. *Cité p.* 46

[Poles 09] D. Poles. *Base of Aircraft Data (BADA) Aircraft Performance Modelling Report.* Eurocontrol Experimental Center, 2009. *Cité p.* 24, 25, 27

[Poles 10a] D. Poles. *Revision of atmosphere model in bada aircraft performance model.* Rapport technique, 2010. *Cité p.* 21

[Poles 10b] D. Poles, A. Nuic et V. Mouillet. *Advanced aircraft performance modeling for ATM : Analysis of BADA model capabilities.* Proceedings of the 29th IEEE/AIAA Digital Avionics Systems Conference (DASC), Oct 2010. *Cité p.* 24

[Ragozin 83] D. L. Ragozin. *Error bounds for derivative estimates based on spline smoothing of exact or noisy data.* Journal of Approximation Theory, vol. 37, no. 4, pages 335 – 355, 1983. *Cité p.* 67

[Ramsay 97] J. O. Ramsay, N. Heckman et B. W. Silverman. *Spline smoothing with model-based penalties.* Behavior Research Methods, Instruments, Computers, vol. 29, no. 1, pages 99–106, 1997. *Cité p.* 66

[Rao 99] C. R. Rao et H. Toutenburg. Linear Models : Least Squares and Alternatives (Springer Series in Statistics). Springer, juillet 1999. *Cité p.* 46

[Reinsch 67] C. H. Reinsch. *Smoothing by spline functions.* Numerische mathematik, vol. 10, no. 3, pages 177–183, 1967. *Cité p.* 66

[Ridgeway 07] G. Ridgeway. *Generalized Boosted Models : A guide to the gbm package.* Update, vol. 1, page 1, 2007. *Cité p.* 57

[Ripley 07] B. D. Ripley. Pattern recognition and neural networks. Cambridge university press, 2007. *Cité p.* 52

[Rissanen 78] J. Rissanen. *Modeling By Shortest Data Description.* Automatica, vol. 14, pages 465–471, 1978. *Cité p.* 42

[RTCA 95] RTCA. *Report of the RTCA Board of Directors' Select Committee on Free Flight.* Rapport technique, janvier 1995. *Cité p.* 12

[Rumelhart 86] D. E. Rumelhart, G. E. Hinton et R. J. Williams. *Parallel distributed processing : explorations in the microstructure of cognition, vol. 1.* chapitre Learning internal representations by error propagation, pages 318–362. MIT Press, Cambridge, MA, USA, 1986. *Cité p.* 52

[SCA 06] Service de l'Information Aéronautique. *Annexe 2 à l'arrêté du 3 mars 2006 modifié (Services de la circulation aérienne)*, décembre 2006. *Cité p.* 7

[Schapire 90] R. E. Schapire. *The Strength of Weak Learnability.* Machine Learning, 1990. *Cité p.* 54

[Schultz 12] C. Schultz, D. Thipphavong et H. Erzberger. *Adaptive Trajectory Prediction Algorithm for Climbing Flights.* AIAA Guidance, Navigation, and Control (GNC) Conference, August 2012. *Cité p.* 2, 33, 92, 95, 96, 97, 185

[Schuster 12] W. Schuster, M. Porretta et W. Ochieng. *High-accuracy four-dimensional trajectory prediction for civil aircraft.* Aeronautical Journal, vol. 116, no. 1175, pages 45–66, 2012. *Cité p.* 33

[Schwarz 78] G. Schwarz. *Estimating the dimension of a model.* The annals of statistics, vol. 6, pages 461–464, 1978. *Cité p.* 42

[SESAR Consortium 07] SESAR Consortium. *Milestone Deliverable D3 : The ATM Target Concept.* Rapport technique, 2007. *Cité p.* 1, 12

[Shanno 70] D. F. Shanno. *Conditioning of Quasi-Newton Methods for Function Minimization.* Mathematics of Computation, vol. 24, no. 111, pages pp. 647–656, 1970. *Cité p.* 52, 142

[Sietsma 91] J. Sietsma et R. J. F. Dow. *Creating artificial neural networks that generalize.* Neural Networks, vol. 4, no. 1, pages 67 – 79, 1991. *Cité p.* 53

[Silverman 85] B. W. Silverman. *Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting.* Journal of the Royal Statistical Society. Series B (Methodological), vol. 47, no. 1, pages pp. 1–52, 1985. *Cité p.* 67

[Simpkin 10] A. Simpkin. *Derivative Estimation in Noisy Data ; an Additive Penalty P-Spline Approach.* PhD thesis, Dec 2010. *Cité p.* 66

[Slater 02] G. L. Slater. *Adaptive improvement of aircraft climb performance for air traffic control applications.* Proceedings of the 2002 IEEE International Symposium on Intelligent Control. IEEE conference publications, October 2002. *Cité p.* 92

[Slattery 97] R. Slattery et Y. Zhao. *Trajectory synthesis for air traffic automation*. Journal of Guidance, Control, and Dynamics, vol. 20, no. 2, pages 232–238, 1997. *Cité p.* 33

[STATFOR 13] STATFOR. *EUROCONTROL 20-year Forecast of Annual Number of IFR Flights (2012 -2035)*. Rapport technique, Eurocontrol – Air Traffic Statistics and Forecasts, Brussels, juin 2013. *Cité p.* 1, 7, 12

[Stone 74] M. Stone. *Cross-Validatory Choice and Assessment of Statistical Predictions*. Journal of the Royal Statistical Society. Series B (Methodological), vol. 36, no. 2, pages 111–147, 1974. *Cité p.* 66

[Storn 97] R. Storn et K. Price. *Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces*. J. of Global Optimization, vol. 11, no. 4, pages 341–359, dec 1997. *Cité p.* 52, 199

[Sugiyama 07] M. Sugiyama, M. Krauledat et K.-R. Müller. *Covariate shift adaptation by importance weighted cross validation*. The Journal of Machine Learning Research, vol. 8, pages 985–1005, 2007. *Cité p.* 182

[Swenson 06] H. Swenson, R. Barhydt et M. Landis. *Next Generation Air Transportation System (NGATS) Air Traffic Management (ATM)-Airspace Project*. Rapport technique, National Aeronautics and Space Administration, 2006. *Cité p.* 1, 12

[Tastambekov 14] K. Tastambekov, S. Puechmorel, D. Delahaye et C. Rabut. *Aircraft trajectory forecasting using local functional regression in Sobolev space*. Transportation Research Part C : Emerging Technologies, vol. 39, no. 0, pages 1 – 22, 2014. *Cité p.* 2, 3, 17

[Thipphavong 13a] D. P. Thipphavong. chapitre Reducing Aircraft Climb Trajectory Prediction Errors with Top-of-Climb Data. Guidance, Navigation, and Control and Co-located Conferences. American Institute of Aeronautics and Astronautics, Aug 2013. 0. *Cité p.* 17

[Thipphavong 13b] D. P. Thipphavong, C. A. Schultz, A. G. Lee et S. H. Chan. *Adaptive Algorithm to Improve Trajectory Prediction Accuracy of Climbing Aircraft*. Journal of Guidance Control Dynamics, vol. 36, pages 15–24, janvier 2013. *Cité p.* 92

[Tibshirani 94] R. Tibshirani. *Regression Shrinkage and Selection Via the Lasso*. Journal of the Royal Statistical Society, Series B, vol. 58, pages 267–288, 1994. *Cité p.* 46

[Trefethen 97] D. Trefethen L. N .and Bau III. Numerical linear algebra, volume 50. Siam, 1997. *Cité p.* 50

[Valiant 84] L. G. Valiant. *A Theory of the Learnable*, 1984. *Cité p.* 54

[Van Hentenryck 97] P. Van Hentenryck. *Numerica : a modeling language for global optimization.* Proceedings of the Fifteenth international joint conference on Artifical intelligence - Volume 2, IJCAI'97, pages 1642–1647, 1997. *Cité p.* 101

[Vapnik 91] V. N. Vapnik et A. Y. Chervonenkis. *The necessary and sufficient conditions for consistency of the method of empirical risk minimization.* Pattern Recogn. Image Anal., vol. 1, no. 3, pages 284–305, 1991. *Cité p.* 38

[Vapnik 95] V. N. Vapnik. The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA, 1995. *Cité p.* 38

[Vapnik 99] V.N. Vapnik. *An overview of statistical learning theory.* Neural Networks, IEEE Transactions on, vol. 10, no. 5, pages 988–999, Sep 1999. *Cité p.* 36, 38

[Warren 98] A. W. Warren et Y. S. Ebrahimi. *Vertical path trajectory prediction for next generation ATM.* Digital Avionics Systems Conference, 1998. Proceedings., 17th DASC. The AIAA/IEEE/SAE, volume 2, pages F11/1 –F11/8 vol.2, oct-7 nov 1998. *Cité p.* 92

[Warren 00] A. W. Warren. *Trajectory Prediction Concepts for Next Generation Air Traffic Management.* 3rd USA/Europe ATM R&D Seminar, June 2000. *Cité p.* 92, 97

[Wilkinson 94] J. H. Wilkinson. Rounding errors in algebraic processes. Dover Publications, Incorporated, 1994. *Cité p.* 99

[Xue 11] M. Xue et H. Erzberger. *Improvement of Trajectory Synthesizer for Efficient Descent Advisor.* AIAA-2011-7020, 11th American Institute of Aeronautics and Astronautics (AIAA) Aviation Technology, Integration, and Operations (ATIO) Conference, Virginia Beach, VA, 20-22 Sep, 2011. *Cité p.* 33

[Zhao 96] Y. Zhao et R. Slattery. *Capture conditions for merging trajectory segments to model realistic aircraft descents.* Journal of guidance, control, and dynamics, vol. 19, no. 2, pages 453–460, 1996. *Cité p.* 33